THE UNIVERSITY *of* EDINBURGH

Physiological and Mechanical Influences on Muscle Function Following Total Knee Arthroplasty

David F. Hamilton

Doctor of Philosophy

University of Edinburgh

2011

Abstract

End-stage osteoarthritis is characterised by pain and reduced physical function, for which total knee arthroplasty (TKA) is recognised to be a highly effective procedure. Post-operative outcome and resultant function however is variable. Many factors are thought to influence outcome; in particular quadriceps muscle strength is one of the strongest predictors of the patient's ability to perform functional tasks. Muscle atrophy has been shown to account for only a third of the variance in muscle power, the remainder is currently unexplained. In this thesis it is hypothesised that physiological and mechanical factors will affect muscle power post TKA.

A new design of prosthesis with an axis of rotation of the knee based on new kinematical observations has been suggested to confer a mechanical advantage to the knee extensor mechanism by lengthening its moment arm, and thus reducing the muscular effort required to extend the knee, however this has not as yet been clinically demonstrated. A strong extensor mechanism is recognised as being paramount to the patients return to functional activity following TKA, but there has been no consideration as to the mechanisms how and to what extent the muscle tissue actually recovers. It is known that muscle satellite cells are essential for the regeneration of skeletal muscle and that these cells are activated following damage, but these have not been considered in relation to recovery from orthopaedic procedures. It is hypothesised that the number of satellite cells in the extensor mechanism will vary in the patient population and will influence muscle recovery.

A double blind randomised controlled trial of 212 TKA patients was conducted to compare the new implant design with a traditional model. Patient outcome was assessed at four points over a one year period. The new implant was superior in measures of knee flexion, lower limb power output and by patient report questionnaire (Oxford Knee Score) Two-way ANOVA, $p = <0.001$ in all cases. Extensor mechanism power was significantly increased between all four assessment

points in the new implant group, the control group demonstrating change between the second and third assessment only ($p = <0.001$).

Analysis of the outcome assessments used demonstrated a changing relationship between function and patient report of that function. Regression models demonstrated that patient report of function became more consistent with direct functional assessment as the influence of pain diminished post-operatively. A hierarchical model is presented that highlights the limitation of patient report data in isolation.

Muscle satellite cells were isolated from biopsies of the quadriceps muscle of 18 patients at the time of surgery and counted by an immunofluorescent staining technique. The number of satellite cells detected accounted for a third of the post-operative variance in power output ($R^2 = 36.6\%$). This was confirmed in another cohort of 11 patients with a more sensitive qPCR technique. It was further found that the activated satellite cells accounted for around two thirds of the change in post-operative power output ($R^2 = 66.7\%$).

In conclusion, both mechanical and physiological factors have a significant effect on muscle power post total knee arthroplasty.

Declaration

I hereby declare that the work described within this document is my own, except where explicitly stated otherwise. This thesis has been composed by me alone, and the material has not been submitted elsewhere as part of other degree or professional qualifications.

David Hamilton
March 2011

Acknowledgements

There are many people who in various ways have contributed to this body of work and deserve my sincere gratitude.

I have been exceedingly fortunate in my supervisory team who have expertly complemented the two arms of this thesis. It is through their active input at professional, academic and indeed personal levels that this work has been made possible.

Prof. Hamish Simpson merits particular thanks for his general oversight of this thesis, continual pertinent advice, considerable insight into the muscle stem cell aspects and the far from inconsequential issue of arranging the funding required to bank-roll this endeavour. I have benefited from funding from both the Medical Research Council doctoral training scheme and from Stryker^{UK}; I extend my thanks to both for their financial assistance.

I am also very much indebted to Mr Paul Gaston for his general support, friendship, and refreshing application of shear common sense. It is through his expertise in all things arthroplasty and considerable capacity to make complex technical issues comprehensible that I have been able to gain a meaningful understanding of the detailed technical and surgical influences that were key to the development of the RCT and indeed this thesis. I further acknowledge his considerable involvement with the development of the implant trial, and practical contribution throughout. Additionally, from a clinical perspective, the many hours we have shared post-operative clinics has directly resulted in an exponential development of my orthopaedic patient management, for which I also thank him.

Further thanks must go to the other consultants and wider arthroplasty team at RIE who have helped facilitate the smooth(ish) running of the clinical trial. Mr Sam

Patton, Mr Richard Burnett, and Mr Colin Howie merit particular thanks and also Mr Gaston's secretaries, Helen and latterly Cathy.

Jennifer McLeish at the Queens Medical Research Centre has donated much of her time, and I thank her for technical assistance with the stem cells, development of the primers used, tireless explanation of the murky depths of qPCR, and that reassuring 'don't worry, this always happens' look while I attempted to optimise the staining protocols.

Thanks also to the members of the Orthopaedic Engineering research group of which I have been a part for the last few years for their general support, coffee and crosswords (which despite their best efforts remain cryptic). Specific thanks to Debbie MacDonald for her expert assistance with the ethics applications, pertinent advice as regards running the clinical trial and in typesetting of this document. Also thanks to Robin Henderson for teaching me the finer points of stepwise regression modelling and to Rob Wallace for numerous Excel shortcuts and the occasional whiskey (with a spoon).

Finally, and most importantly, I thank my family for putting up with all the hours I have spent on this project. Thesis write-ups are notorious for their perpetual nature, and despite my best efforts, it has required a certain selfishness to document this project. My apologies to my wife Anna for the exquisitely poor timing of the write-up phase, which as a result has quite literally left her holding the baby, it is in no way adequate compensation that I dedicate this work to her.

Contents

Table of Figures

1 Introduction, Scope and Outline

1.1 Introduction

The National Collaborating Centre for Chronic Conditions (NCC-CC) describes osteoarthritis as the most common disease of the joints, and one of the leading causes of pain and disability worldwide. The disease is responsible for considerable morbidity (NCC-CC, 2008) and it has recently been identified that patients with osteoarthritis have an excess mortality compared with the general population (Nuesch et al, 2011).

Many therapeutic options are suggested as treatments of osteoarthritis; however the only intervention that demonstrates a large effect size in relieving chronic knee pain is total knee arthroplasty (Juni et al, 2006). The prosthetic implants are known to survive well and have low complication rates (U.S. National Institute for Health, consensus statement, 2003; British Orthopaedic Association / British Association for Surgery of the Knee position statement, 2010; Ethgen et al, 2004). Last year, in the UK alone, over 70,000 total knee replacements (TKA) were performed (National Joint Registry, accessed 2011) while the surgical volume has been projected to increase dramatically in the next 20 years. An American projection of TKA surgical volume is for demand to grow by 673% by 2030, which would constitute some 3.5 million procedures every year in the USA alone (Kurtz et al, 2007). Culliford et al (2010) in a review of UK procedures over a fifteen year period lend support to this view of dramatic increases in demand for TKA.

Total knee arthroplasty is a very successful procedure, the American National Institute for Health (NIH) position statement (2003) explains that the success of primary TKA in most patients is strongly supported by more than 20 years of follow-up data. That rapid and substantial improvement in the patient's pain, functional

status, and overall health-related quality of life is found in 90 percent of cases, and that around 85 percent of patients are satisfied with the results of surgery. Despite this, post-operative outcome and resultant function is variable (Hawker, 2006; Wylde et al, 2007).

It is thought that factors relating to the surgeon (surgical volume and operating institution) inter-operative factors and perhaps post-operative rehabilitation affect subsequent patient outcome (Dennis et al, 2007; NIH consensus statement, 2003). Factors relating to the patient such as general physical condition pre-operatively are also thought to affect resultant post-operative outcome. Pre-operative quadriceps strength in particular is one of the strongest predictors of post-operative function in terms of the patient's ability to perform functional tasks (Faulkner et al, 2008; Lingard et al, 2004; Lamb and Frost, 2003). Pilot work has shown however that muscle atrophy accounts for around only a third of this variance, the rest, in the short term, has been explained by failure of muscle activation following surgery (Mizner et al, 2005). Long term strength deficits though are well reported compared to the patients opposite limb or healthy controls (Huang et al, 1996; Silva et al, 2003). This discrepancy has not been explained.

Developments in implant design have striven to improve patient outcome through a variety of technical changes, though the British Orthopaedic Association (BOA) / British Association for Surgery of the Knee (BASK) position statement on good practice in TKA (2010) notes that it is debatable whether there has been any meaningful improvement in prosthesis design since the blueprint of all modern 'condylar' knee implant designs in the 1970s.

Recently however a new theory of the underlying kinematic motion of the knee has emerged, that challenges the axis of rotation that all traditional knee implants are set to recreate (Hollister et al, 1993; Churchill et al, 1998). A new single-radius of curvature design based on this new kinematic theory may substantially improve the patient outcome by more closely matching the 'normal' kinematic motion of the knee

and as a result mechanically lengthening the moment arm of the extensor mechanism effectively reducing the effort needed from the quadriceps muscle to extend the knee (Hall et al, 2008; D'Lima et al, 2007; Wang et al, 2006). This theoretical advantage has not to date been clinically demonstrated.

The mechanisms involved in the regulation of skeletal muscle growth and regeneration are of great scientific and clinical interest, as it is believed that therapeutic manipulation of these mechanisms can improve the quality of life of individuals suffering from a plethora of conditions as diverse as muscular dystrophy, chronic heart failure and indeed sarcopenia (Spangenberg and Booth, 2001).

It is known that muscle satellite cells (myogenic precursor cells) are essential for the regeneration of skeletal muscle and that these cells are activated following damage; this process has been well described (Collins, 2005; Hawke and Garry, 2001). In the future it may be possible to regulate the proliferation of the satellite cells, via gene delivery to the skeletal muscle. We may further be able to isolate the satellite cells, genetically manipulate them and deliver them back to the muscle via the circulation (Spangenberg and Booth, 2001), to enhance muscle recovery and patient well being. This however will be dependant on substantial further advances in biological technology.

In the field of orthopaedics improved regeneration of muscle would have direct benefits as enhanced patient outcome following surgery is gaining increasing prominence, yet to date this has not been investigated. While there may be a general application of potential muscle therapy techniques, for example to patients undergoing leg lengthening procedures, the largest single population who potentially stand to benefit are those suitable for total knee arthroplasty. In addition to the potential volume, knee arthroplasty is possibly the most relevant operation to consider due to the importance of the extensor mechanism, and specifically the quadriceps muscle group, in subsequent rehabilitation and eventual outcome.

A strong extensor mechanism is recognised as being paramount to the patients return to functional activities following surgery (Mizner, 2005b). There appears however, to be no consideration as to the mechanisms how and to what extent the muscle tissue actually recovers post-operatively.

1.2 Scope of thesis and research hypothesis

This thesis is concerned with two poorly explored influences on post-operative muscle performance within the context of patient outcome following total knee arthroplasty; (1) the role of prosthetic design that confers a mechanical advantage to the quadriceps and (2) the role of the patient's underlying regenerative capacity of the quadriceps.

The aim of this thesis is to determine whether either the mechanical factor of prosthetic design or the physiological factor of regenerative capacity of the patient's muscle influences the patient's physical function following total knee arthroplasty.

The primary hypothesis is that both mechanical and physiological variables will influence the ability of the patient to generate lower limb power post operatively, and thus benefit overall patient function.

Three specific research questions address this hypothesis:

1. Does implant design that mechanically advantages the musculature of the extensor mechanism enhance muscle power and physical outcome following TKA?
2. Does the number and activation state of the quadriceps muscle satellite cells influence the recovery of extensor mechanism muscle power post-operatively?
3. To what extent is physical recovery expressed through standard patient outcome assessments following TKA, and how does the extensor mechanism power output assessment relate to these?

1.3 Outline

The background of total knee arthroplasty, factors that influence patient outcome and the methods of assessing this outcome following surgery are introduced in Chapter 2. The literature concerning single radius implant designs that theoretically enhance the function of the extensor mechanism is critiqued, as is that of the muscle satellite cell. The limited investigation of these cells in human populations is also highlighted.

The clinical evaluation of a new implant based on this differing understanding of the kinematics of the knee is presented in Chapter 3. The new design is compared with that of a standard implant in a double blind randomised controlled trial to assess both overall patient outcome and specific extensor mechanism power output at 1 year. Respective recovery of function within that year is also assessed.

Chapter 4 considers the differing levels of information conferred by different types of outcome assessment, to determine the association between patient reports of function and direct evaluation of this (including power output assessment). A novel aspect is the consideration of the relationship between patient report of outcome and direct measurements at differing time points prior to and following TKA. Regression models are constructed for this relationship and a theoretical assessment framework to explain the interaction of these assessment methods presented.

The influence of the muscle satellite cell is considered in Chapters 5 and 6. Biopsies of the quadriceps muscle are taken and immunohistochemical methods used to locate and count the cells. Data is analysed in the context of muscle power recovery following total knee arthroplasty. The relevance of intrinsic satellite cell number on subsequent post-operative muscle power is compared with the influence of pre-operative power values. More precise molecular biology techniques used to determine the satellite cell content are developed in Chapter 6. These allow the

consideration of the state of activation of the satellite cell, which is further investigated as to the effect on patient lower limb power recovery following TKA.

The conclusions of the thesis are presented in Chapter 7 in addition to suggested methods by which future research could be promoted.

2 Literature Review

2.1 History and Development of Knee Arthroplasty

The concepts of knee arthroplasty and prosthetic implants have changed dramatically over the last century. Whilst this change has been evolutionary in nature, a progression through three broad phases has defined changes in practice. Soft tissue arthroplasty initially dominated thinking before the advent of metal hinges; it was not until the 1970s that the blueprint for all modern condylar designs emerged.

Interpositional Arthroplasty

In 1861 Ferguson reported the first successful soft tissue interposition knee arthroplasty. 5 years post-operatively the patient was described as having a 'useful limb', which established that interposition of soft tissues or foreign material into the knee joint could prevent ankylosis (Shetty et al, 2003a). Baer (1918) published a series of 28 cases of interpositional knee arthroplasties using pigs bladders tanned in potassium chloride, reporting 15 'good' outcomes, though the exact definition of 'good outcome' at that time must be considered different to that accepted today. Around the same time Putti (1920) reported good results interposing the Tensor Facia Lata into the knee. This interpositional technique was used until as recently as 1958 with skin flaps (Brown et al, 1958). Problems such as soft tissue shortening, infection, inflammatory response and re-ankylosis were however associated with this technique, driving research with alternative materials.

Hinges

Cobalt-chrome (Vitallium, Howmedica) was first used as a material in 1938, though gained prominence in the 1950s. It had excellent wear characteristics and was non-corrosive, which constituted a major technological change and subsequently soft tissue interpositional arthroplasty dwindled in popularity. This development has led directly to today's use of titanium alloy, stainless steel 316L and cobalt alloy F75 (Shetty et al, 2003a).

Tibial hemi-arthroplasty had been suggested as early as 1894, when Gluck described the use of an ivory cup with an intramedullary peg inserted into the tibia. The concept of tibial (and femoral) hemi-arthroplasty didn't emerge as a technique until 1950 using steel implants (Marquardt, 1950). Mackintosh (1966) refined this, using techniques not dissimilar to those employed today: correction of bony deformity and removal of minimal bone and ligaments. Metallic tibial blocks were not fixed but held by tension in the collateral ligaments; an early example of ligament balancing. Different sized blocks were used to correct for varus / valgus deformity (Shetty et al, 2003a). Good results were reported with the elderly, but interestingly not the 'active young'. Also of note is that most early surgery was in rheumatoid and not osteoarthritis patients, the degenerative process having not been properly identified at this time.

Combining tibial and femoral components, joined by a steel rod to create a hinged prosthesis (Walldius, 1953) was the next major development in arthroplasty (Shetty et al, 2003b). These were introduced in an attempt to provide stability via the implant. Initially acrylic was employed then metal due to breakage of the acrylic (Walldius 1957 and 1960). This simple hinge allowed a theoretical 115 degrees of flexion. Jones (1969) reported a series of 80 Walldius replacements, commenting 75% were pain free on weight bearing, and 85% exhibited over 60 degrees of flexion. These results were tempered by high failure rates due to aseptic loosening

and infection (requiring removal and arthrodesis) and that patient selection favoured those who were older with limited walking ability. Semi-constrained hinges were developed to allow tibial rotation; these also had high failure rates and were bulky and complex. Shetty et al (2003b) comment that a failure to understand that instability was actually due to the cartilage and bone loss and not the disruption of the collateral ligaments delayed the introduction of reliable joint replacements for a further twenty years.

Condylar replacements

Gunston (1971) suggested the complexity of the knee joint could not be recreated by a simple hinge. Movement was not in a single axis (as the hinge) but constituted a roll and glide of the femur on the tibia: a multi instant centre of rotation. He also introduced the combination of metal, plastic and cement, performing articular surface replacements. Twenty of 22 procedures were reported as good at two years (Gunston, 1971). This was important as it demonstrated that instability was due to cartilage and bone loss, as opposed to collateral ligament destruction (Shetty et al, 2003c), this minimalistic implant actually facilitating better physiological stability than the large hinges.

Freeman and Swanson introduced a similar implant in 1972, and the tensor instrument in 1974 that is still used today for ligamentous gap balancing to control varus / valgus alignment (Todd et al, 1978). The Freeman prosthesis (the ICLH) was the first to be subjected to a large multi-centre trial. Gibbs et al (1979) reviewed 75 prostheses and reported 80% success (defined as minimal or no pain, flexion in excess of 60 degrees and no major complications). Insall, Ranawat and Walker (later with Burnstein) developed their similar design of total condylar replacement and were the first to introduce the tri-compartmental replacement, the lineage of which is used today. The total condylar replacement (Insall et al, 1976) has set the standard of modern implants, with 94% survivorship at 15 years (Ranawat et al 1993).

Modern developments

Most modern knee implants are based on the condylar design, and achieve excellent rates of survival (NIH consensus statement, 2003; Liow and Murray, 1997: Pradhan, 2006). The American Academy of Orthopaedic Surgeons estimates that over 150 different knee implants are currently available for use (AAOS, 2010).

The BOA/BASK position statement on good practice in TKA (2010) states that the selection of knee prostheses for general use should be based on evidence published in peer reviewed journals with a clinical follow-up of at least 10 years and calculated survival curves (Liow and Murray, 1997) demonstrating 90% implant survivorship at 10 years. In the absence of this, devices should be subject to ongoing surveillance and be part of properly controlled prospective trials (BOA/BASK, 2010). Published results of many knee implants, however, offer little help to the surgeon wishing to make an informed choice, as most outcome research is short term, non-comparative and doesn't take into account variations in operating technique (BOA/BASK 2010).

Jacobs et al, writing in a Cochrane Collaboration Review, comment that to improve on an already successful procedure it is the small details that require investigation (Jacobs et al 2007a). Many developments in implant design have been aimed at improving quality of life, and the duration of prosthetic survival (KAT trial group, 2009), though there is little consensus as to whether some recent design changes genuinely affect patient outcome.

Many factors influence the surgical and functional outcome of TKA, and choice of prosthesis may have an important influence (NIH, 2003). Most of the current developments in design have failed to produce substantive benefits in terms of patient outcome. An example of this is the effect of bearing design, where mobile designs have theoretical advantages compared to fixed systems but these have not

been established biomechanically or clinically (NIH position statement, 2003; Wylde, 2007; Huang, 2007).

Mobile designs were introduced with the intention of reducing polyethylene wear and component loosening, as the mobility of the tibial surface allows both low contact stress and constraint forces (Sathasivam and Walker, 1994; McEwan et al, 2005). Huang et al (2007) in a review article note equivalent survival rates of around 95% at 10 years for both designs and equivalent patient outcome (Range of motion, patellofemoral complication rates, American Knee Society Score) in comparative studies. A Cochrane review (Jacobs et al, 2007b) described the overall quality of research methodology in this area as poor, as only one suitable randomised trial (Price et al, 2003) could be analysed. These authors demonstrate superior early results with mobile bearing components, however it is conceded that that the difference cannot be solely linked to the bearing surface as different prostheses types were used in the study. Since this review, two multi centre randomised trials (Wylde et al, 2007; KAT trial group, 2009) have found no early difference in outcome at to two years post-operation.

Other current examples of design changes that have been made to improve outcome, but for which clinical evidence is lacking are metal backed tibial components (KAT trial group, 2009), patella resurfacing (Pakos et al, 2005; KAT trial group 2009), the role of high flexion implant designs (Nutton et al, 2008) and the retention/sacrifice of the posterior cruciate ligament (Jacobs et al, 2007a). The four bar linkage system explains that the anterior cruciate ligament is of key kinematic importance in the knee (O'Connor, 1989), and crucial to the functioning of the successful unicompartmental knee arthroplasty procedure. It does not have the same influence in primary total condylar knee arthroplasty, where this ligament is sacrificed.

2.2 Knee Arthroplasty Outcome

Success of total knee arthroplasty

Whilst generally a very successful procedure, patient outcome following TKA is not universally positive, as a significant number of patients continue to experience pain and functional disability post surgery (Hawker, 2006).

The American National Institute for Health notes that short-term outcomes were generally substantially improved following TKA across the spectrum of presenting disability (NIH position statement, 2003). Despite this, unimpaired function is rare post-operatively, and some degree of limitation should be expected. Although clearly a very effective procedure in the majority of cases, some patients receive no benefit (Wylde et al, 2007). Chronic pain is the major reason that people elect to undergo TKA, thus pain relief is a key outcome post-operatively. Wylde et al (2007) notes that the majority of literature suggests that TKA provides good pain relief, and that the majority of patients experience no (or very mild) pain, however a subsection experience moderate to severe levels.

Technical factors (including surgical technique and alignment of the prosthesis) have been noted to influence both short and long-term outcomes, while factors related to a surgeon's case volume, and choice of prosthesis also have important influences (NIH, 2003; Dennis et al, 2007). Peri-operative complications, post-operative rehabilitation and patient factors are also accepted as affecting outcome (Dennis et al, 2007). Not all differences in functional outcome can be accounted for by these surgical variables, those that remain are, as yet, medically unexplained (Wylde et al, 2007). Many risk factors for poorer outcome have been identified, and include socio-demographic, psycho-social and medical factors.

Lingard et al (2004) in a multi-centre study of 860 patients highlight the role of pre-operative status. While the poorest functioning patients pre-operatively make the largest proportional gains, they retain the lower spectrum of post-operative functional scores. Low pre-op physical and mental health scores are the strongest determinants of limitation to post-operative function at 1 and 2 years. Franklin et al (2010) also describe substantial variation in function post TKA, despite 'excellence in surgical technique and consistent pain relief'. In a review of over 8000 patients, they also found pre-operative variables to be associated with poorer improvement in post-operative function. Interestingly, they reported a bimodal distribution in post-operative SF-12 physical component scores (noting a normal distribution pre-operatively), thus suggesting two distinct groups of responders and non-responders. A higher chance of 'less functional gain' was reported for patients with a BMI over 40, low mental health scores, increasing age and poor quadriceps strength. The BMI and poor quadriceps strength had the most impact, with more than a 2-1 odds ratio of poor functional gain post-op.

In contrast, Bourne et al (2007) reported a consecutive series of 728 patients from a single centre, all with the same implant, thus controlling for many of the known surgical variables. Similar outcomes were described when change in outcome score from pre-op to post-op was assessed regardless of gender, age or obesity class. They suggested that this was a more appropriate analysis. A report from the Edinburgh unit, in a paper comparing outcome following hip and knee replacement, found that neither age nor gender are significant in predictive regression models of 1 year outcome (Hamilton et al. manuscript under review, after modifications requested by the editor, Appendix A). A consecutive series of 1244 knee replacements and 1410 hip replacements from this centre were reported. Only two factors were predictive of change in Oxford Scores; the type of surgery (THA or TKA) and the pre-operative Oxford Score. This same study highlighted significant differences in physical outcome between arthroplasty groups. The knee arthroplasty group demonstrated worse levels of physical function and twice the rate of dissatisfaction (20% dissatisfied) 1 year post-operation.

Another potential variable is the trend to operate on younger patients with increased expectations of high levels of post-operative activity. Mont et al, (2008) report similar clinical and radiographic outcomes in groups of active and sedentary patients 7 years post TKA. The active group undertook many of the activities suggested as suitable post TKA (swimming, cycling and doubles tennis), however around 20 percent of this group were able to participate in higher stress sports (such as jogging, singles tennis, football and basketball). The contrast between the reports of high levels of function (Mont et al, 2008) and those of cohorts of poorly functioning TKAs (Hawker, 2006) suggests the potential for a vast discrepancy to exist between the upper and lower ability levels of patients following knee arthroplasty. These highly functioning individuals may be masking the poor responders when average population scores are reported; this is particularly relevant if change in score is not assessed.

Assessment of outcome

Orthopaedic interventions are assessed on their outcome, and outcome scores are now commonly used in the TKA research literature to compare prostheses, surgical techniques, post operative care and also in auditing departments and individual surgeons (Beard et al, 2010; Ashby et al, 2008).

Survival curves are the standard long term evaluation of implant longevity, for which revision procedure is generally accepted as the end-point (BOA/BASK, 2010; Liow and Murray, 1997). This analysis is criticised as not adequately defining the failure of the implant, nor commenting on patient function prior to the diagnosis of failure (Price et al, 2010). Previously implant designs could be differentiated as to their respective merit by their failure rates. However the advent of modern implants and enhanced surgical procedures has devalued this assessment, as most achieve comparable results.

Assessment of outcome is shifting away from the dichotomous criteria of success or failure (Ashby et al, 2008). The combination of comparable implant survival, continual prosthetic design development, changing functional expectation post surgery and the understanding that some patients do not benefit from TKA has driven the need for reliable short term information as to the success of surgery. The use of patient reported outcome measures (PROMs) to assess functional outcome and quality of life following TKA is advocated as additional information (Bream et al 2010; Beard et al, 2010; Price et al, 2010; Bream and Black 2009), and recently has been adopted as a requirement of all providers of elective arthroplasty surgery of NHS patients (DoH, 2008).

Objective clinical assessment of outcome following TKA is routinely performed by the surgical team, though formal scoring systems that are assessor-based suffer observer bias, and have been criticised for this (Beard et al, 2010). Patients are suggested to offer a complimentary perspective to that of the clinician into the effectiveness of health care. Clinicians can make objective observations as to impairment and disability, but only patients can report on their quality of life (Black and Jenkinson, 2009). Advocates of PROMs say that they provide a remarkably sophisticated measure of whether a treatment has worked in the (important) sense of whether or not the patient feels better, and how much better (Timmins, 2009).

Numerous scales exist to determine post-operative function (Beard et al, 2010; Riddle et al, 2008), though no gold standard tool for assessing outcome following joint arthroplasty has been developed (Davies, 2002). The current consensus statement from the British Orthopaedic Association and the British Association for Surgery of the Knee (2010) states that there is no agreed standardisation of outcome measures for TKA. Many trials report similar research questions but with differing outcome measures which makes it difficult or often impossible to compare the trials or carry out systematic reviews (Riddle et al, 2008; Davies, 2002). Beard et al (2010) note that many of these scoring systems do not meet the recommended quality criteria around validity, reliability and responsiveness; further complicating the issue.

Broadly PROMs questionnaires fall into three categories; generic health assessments, disease specific instruments and joint specific tools (Ashby et al, 2008). These reflect the suggestion of the OMERCAT group, that three core patient reported domains related to the symptomatic severity should be evaluated: pain, physical function and patient global assessment (Bellamy et al, 1997). A combination of assessment tools may then form the best outcome analysis (Beard et al, 2010).

Performance measures

A disadvantage of patient report based scales for assessing post-operative function is that they measure the patient's perception of their knee status and are thus subjective (Stratford and Kennedy, 2006). As a consequence they are thought to be influenced by pain (Terwee et al, 2006; Stratford and Kennedy, 2006), which results in a low functional content validity (Boonstraa et al, 2008). High levels of content validity are required to evaluate biomechanical aspects of function, and generally, performance based measures demonstrate this (Boonstraa et al, 2008; McCarthy and Oldham, 2004).

It has been suggested that self report questionnaires and performance measures may assess different aspects of physical function. Stratford and Kennedy (2006) suggest that in terms of expressing the difficulty patients have in moving around or looking after themselves, self-report measures provided information concerning the experience associated with doing the task, while performance measures confer information about the ability to do the task.

The Outcomes Measures in Rheumatology Conference specified that patient report measures of functional outcome are recommended for all phase III clinical trials; whereas performance based measures of physical function are optional (Bellamy et al, 1997).

PROMs measures are commonly used as they are comparatively cheap, effective at collecting large volumes of data, and do not require follow-up clinic visits to achieve this (Mizner et al, 2011 published ahead of print). Woods et al (2008) however justified the relevance of physical assessment tools as offering additional insights into clinical progression and prognosis. The suggested discrepancy between patient report questionnaire assessment of function and objective measure of patient performance (Beard et al, 2010; Stratford and Kennedy, 2006) suggests that both assessments should be used in tandem to provide the most complete assessment (Witvrouw et al, 2002). One assessment type should perhaps not be considered preferential to the other, as they provide distinct but complementary information. This is akin to the suggestion that generic and specific PROMs should be employed. The choice of assessment tools depends on a variety of factors relating to the aim of the study, level of detail of observation required, funding and the context of the study (Beard et al, 2010; Ashby et al, 2008).

2.3 Muscle Power and Functional Outcome

Green and Schurman (2008) comment that post-operative muscle strength is a critical part of the success of TKA, and further state that insufficient quadriceps strength is recognised as a relative contraindication for the procedure. Adequate function of the extensor mechanism is essential to allow participation in most activities of daily living (Noble et al, 2006; Silva et al, 2003), and knee extensor weakness has been reported to be closely associated with mobility limitations in the TKA population, particularly when evaluating functional tasks such as climbing stairs and transferring from a chair (Mizner et al, 2005c; Valtonen et al, 2009). Several studies have shown difficulties in both walking (Yoshida et al, 2007; Walsh et al, 1998) and stairs ascent (Walsh et al, 1998) post TKA compared to controls. Relatively little data however exist on knee strength post TKA and its relationship to patient characteristics and alternative outcome measures (Silva et al, 2003).

Power is better related to function than measures of absolute strength due to the role of velocity of movement in every day tasks (Bassey and Short, 1990). Several investigators note lower limb power deficits in excess of 20% compared to the non-operated knee up to six months post TKA (Gapeyeva et al, 2007; Berman et al 1991; Mizner et al, 2005a; Lorentzen et al, 1999). This deficit would appear to continue over time, though the studies reporting this suffer from poor methodological design.

Valtonen et al (2009) assessed the power output of 50 unilateral TKA patients at an average of 10 months post-op. They found mean knee flexion power deficits of 19% and extension power deficits of 23% compared to the un-operated limb. Silva et al (2003) demonstrate 30% deficits compared to healthy controls at 3 years post-op. Both investigations suffer from a lack of pre-operative data. There is only a single paper that reports longer term power data (Huang et al, 1996); they noted deficits in hamstring to quadriceps peak torque ratio of TKA patients compared to controls at 6-13 years post-op. This paper reported the results of 50 replacement knees in 34 patients and 16 control knees in 9 patients. Unfortunately, the reporting of multiple joints per individual, a lack of pre-operative data or stratification by subject characteristics, prevents comparison of absolute strength between patients and controls. Berman et al (1991), in contrast, noted a normalisation of flexion extension ratios, compared to the contralateral uninvolved limb by 2 years post TKA in a well conducted prospective study involving 68 patients. Interestingly the hamstring peak torque reached the levels of the uninvolved knee between 7 and 12 months post-op, while the quads mechanism continued to demonstrate slightly reduced peak torque at 2 years post-op, suggesting specific limitations in the rehabilitation of this muscle group.

Pre-op function (Lingard et al, 2004) and quadriceps muscle strength (Franklin et al, 2010) have been suggested as the strongest predictors of post op function. Lamb and Frost (2003) assessed the predictors of recovery of mobility at 6 months post TKA in a series of 79 patients. Mobility was measured using timed tests of walking and stairs climbing. Pre-operative variables considered included age, gender, comorbidities,

flexion BMI and lower limb power. Pre-operative power most strongly predicted walking speed and stairs ascent speed at six months (BMI and pain were associated to a lesser extent; no other variables were significantly associated with either measure). Lamb and Frost (2003) is the only report to also assess recovery of mobility. Lower limb power was, by a large margin, the strongest predictor of change in walking speed and change in stairs ascent speed between pre-operative and 6 month assessments (comorbidities and BMI the only other variables associated). Mizner et al (2005c) corroborate this, reporting pre-operative quadriceps strength, but not pain or range of motion to be significant predictors of post of function (stairs climbing or chair transfer) in a study of 40 arthroplasty patients at 1 year.

The assertion that pre-operative flexion does not affect post-operative function is in keeping with the lack of benefit found with high flexion implant designs. Nutton et al (2008) in a well designed RCT investigated the role of an increased posterior femoral offset high flexion design modification compared to the standard implant that it was based on. No clinically relevant differences were found. Interestingly both implant groups in this study used generally the same range of motion to complete daily functional tasks; however the range used by both implant groups was substantially below that of an aged matched control group of healthy individuals.

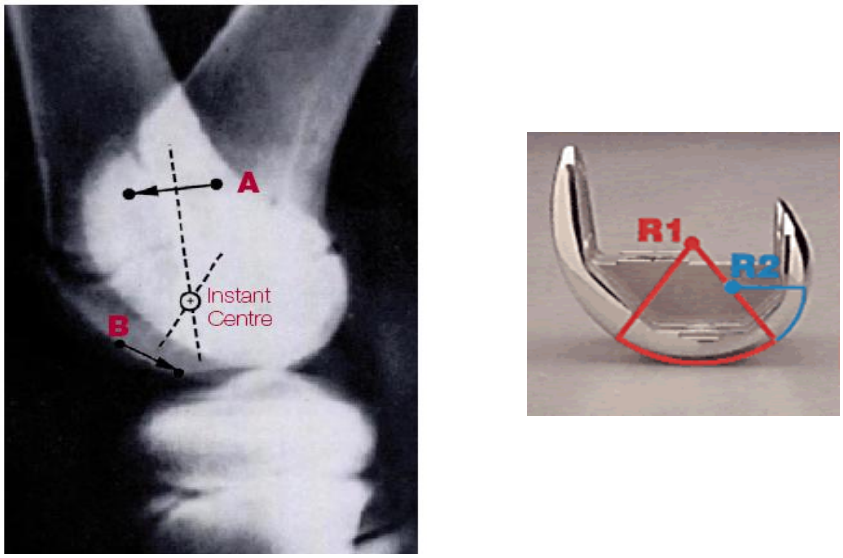
The current literature indicates that the dominant factor in physical recovery post-operatively relates to the extensor mechanism function; however muscle atrophy alone does not explain the variation found. Mizner et al (2005a) suggest that the recovery of function is related not simply to addressing muscle atrophy, but to the failure of voluntary muscle activation. They suggest that failure of voluntary activation contributes twice as much as atrophy to the loss of quadriceps strength in the early post-operative phase. Further that as much as 85% of the loss of quadriceps strength can be explained by these two factors. While an attractive idea, that resonates with clinical physiotherapy practice in terms of post-operative rehabilitation, these conclusions are based on a study of 20 patients. Disappointingly,

these exciting pilot results published in 2005 have yet to be corroborated by a large clinical trial.

2.4 Single Radius Implant Design

Classic theory of the kinematic motion of the knee suggests that flexion and extension occur around a changing instant centre of rotation (Figure 2.1). This conceptual axis is located relatively anterior and proximal in extension moving posterior and distal in flexion. This theory underpins the traditional femoral component design of all multi radius implants (Mahoney et al, 2002).

Figure 2.1 – Images showing the instant multi-centre of rotation implant design. The shifting centre of rotation is highlighted in both the Radiographic image (the Releaux method to show the 'Instantaneous Centre of Rotation' between two flexion angles) and the accompanying diagram which demonstrates the resultant implant design and the changing centre of rotation within.



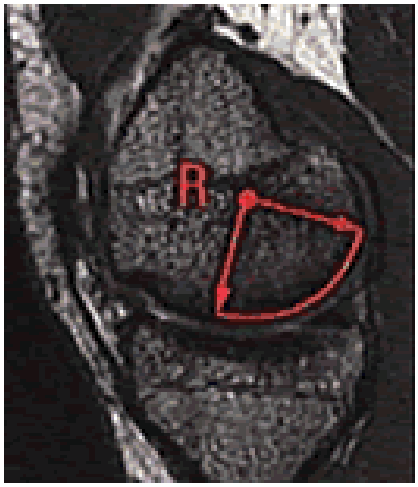
Images adapted from Stryker Orthopaedics instructional publications.

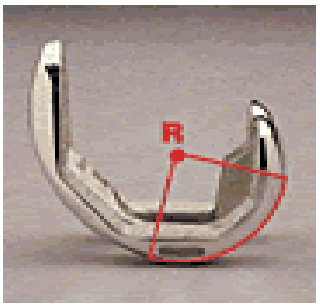
More recently, the validity of the anatomical observations that are the basis of the shifting instant centre of rotation and underpin the multi-radius design have been challenged (Panjabi et al, 1982) and recent anatomical studies have led investigators to suggest that knee flexion / extension occurs around a single fixed axis in the femur

(Hollister et al, 1993) that approximates to the transepicondylar axis (Churchill et al, 1998).

The single radius of curvature design of femoral component (Figure 2.2) is based on the premise that there is a single axis which is fixed on the femur, as opposed to the instantaneous shifting centre of traditional designs which have two or more radii of rotation within the functional range of motion of the knee (Hoshino, 1997).

Figure 2.2 – Images showing the single axis of rotation implant design. Radiograph highlighting the single flexion / extension axis from 10 to 100 degrees of knee flexion. The accompanying image demonstrates the same axis on the resultant implant design.





Images adapted from Stryker Orthopaedics instructional literature.

This difference in axis of rotation establishes a major biomechanical difference in the implant design that results in a more posterior axis of rotation in the femoral condyles (D'lima et al, 2001; Churchill et al, 1998, Hoshino, 1997: Pinskerova et al, 2000). A consequence of this is the moment arm of the knee extensor mechanism is lengthened (Figure 2.3).

Figure 2.3 - Diagrammatic representation of the hypothesised reduction in quadriceps muscle force generation required to extend the knee as a result of the increased moment arm of the single axis of rotation

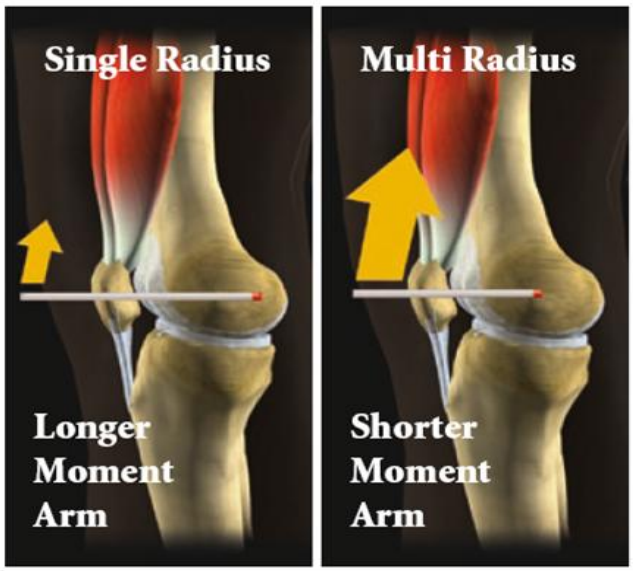


Image adapted from Stryker Orthopaedics instructional literature.

It has long been recognised that increasing the femoral offset of a total hip arthroplasty provides the abductor muscles with a longer moment arm which decreases the muscle force required to abduct the lower limb and lowers the joint reaction force. The same principle can be applied to knee arthroplasty, where a more posterior flexion / extension axis lengthens the knee extensor moment arm, decreasing the quadriceps muscle force required to extend the knee and reducing the joint reaction force (Mahoney et al, 2002). A failure to restore the normal knee extensor moment arm following TKA with the traditional multi radius design of implant has been reported (Huang et al, 1996, Singh and Schmalzried 1996), and the single axis concept provides a convenient explanation as to the reasons for this, and suggests a potential biomechanical benefit of designs based on the single axis theory.

The extensor mechanism is the fundamental dynamic support of the knee during stance and locomotion (Gomez-Barrena et al, 2010). Adequate function of which is a prerequisite for participation in many activities of daily living (Silva et al, 2003), and crucial in promotion of positive clinical outcome and postoperative satisfaction

(Noble et al, 2006). Enhanced muscle recovery based on the single radius femoral components is claimed by the manufacturer (Stryker literature). A cadaver study suggests reduced patello-femoral joint forces with a single radius design (D'Lima et al 2001), but little is known about the consequential effects of single radius design on physical function (Wang et al, 2006).

Clinical studies

Wang et al (2006) investigated the effect of differing radii of curvature implant designs on the patient's ability to stand from a seated position using a complex analysis of video motion tracking and isokinetic dynamometry with complementary (surface) electromyography. The authors note significant biomechanical differences in time to complete the performance task, kinematic values and patterns of electrical activation between TKA groups during the sit-to-stand motion. Specific differences in compensatory mechanisms (of trunk flexion angle and velocity) adopted to achieve the task were observed in the multi radius group that were absent in the single radius group.

Previously Su et al (1998) have noted increased trunk flexion angle in patients following TKA compared to controls to compensate for extensor muscle weakness. A brief summary of the mechanism being that increased trunk angle reduces the extension moment across the hip joint, which subsequently reduces the knee extension moment, thus requiring less knee extensor muscle force.

Wang et al (2006) report a tendency for increased trunk flexion velocity of 7 degrees per second in the multi radius group to generate the standing motion. Theoretically, if increased trunk flexion velocity is demonstrated, an increased deceleration of this motion must also occur. As the hip extensors pull the trunk backwards to reduce the forward momentum, at the same time a transfer of mechanical energy from the trunk to the thigh occurs which acts to rotate the thigh to a vertical position. (Riley et al,

1991), i.e. horizontal momentum of the trunk can be transferred into vertical momentum of the thigh. This compensatory strategy may reduce the knee extensor moment and resultant quadriceps activity needed to achieve the movement. Interestingly Wang et al found increased muscle activity (EMG analysis) in the multi radius group compared to the single radius group (1 way ANOVA $p < 0.05$).

While the results are suggestive of differing performance between single and multi radius design, there is some ambiguity with the data analysis and reporting. The study of Wang et al is a single assessment of 16 patients at least 18 months following TKA. The patients are described only as being from a group of well functioning TKAs in a single surgeon series. Eight single radius implants (Scorpio, Stryker orthopaedics) and eight multi radius implants (Series-7000, Stryker Orthopaedics or PFC, DePuy Orthopaedics – relative numbers are not stated) are reported. The single radius group is younger at 65 (+/-5) years than the multi radius group at 72 (+/-7) years, and assessed much earlier post-op, 29 (+/-11) months compared to 79 (+/- 25) months respectively.

The statistical analysis is somewhat confusing. It is stated that 1-way ANOVAs were compared to 1-way analysis of covariance to assess the role of the age and post-operative time. These covariates were found to affect the time of three variables: standing time, trunk flexion velocity and knee extension velocity. Despite conclusions being drawn on between group differences using these variables, the results reported are for the ANOVAs, not the analysis of covariance, that the authors state affects the significance of the outcome. The authors justified this in their conclusion by stating that the results are still significant when controlled for the covariates, however the details remain elusive and unreported, thus true differences can only be speculated upon. It is also unclear the extent to which these very specific differences in function are clinically relevant.

Three further sets of authors have investigated the clinical outcome of single radius designs in terms of the extensor mechanism efficiency, though all suffer from issues

around their study design. Mahoney et al (2002) reported improved function of the extensor mechanism post TKA and faster restoration of flexion in association with design features that increase the length of the extensor moment arm. Analysis of a consecutive series of patients (from a single surgeon) was reported, with the stated aim to compare the last 100 multi radius TKAs (Series 7000 PPSK, Osteonics, NJ) performed with the first 100 single radius TKAs (Scorpio, Osteonics, NJ). 184 knees in 150 patients were subsequently reported without adequate discussion of the reduced numbers (83 multi radius knees in 74 patients and 101 single radius knees in 76 patients). A further confounding issue is the inclusion of the 34 patients who had bilateral knee implants (9 in the multi radius group, 25 in the single radius) and no discussion as to how this affected the results, beyond commenting that bilateral patients were slower to rehabilitate. If this is the case, then the different proportion bilateral procedures in each group is clearly relevant.

Evaluation consisted of multiple testing time points: pre-operatively and then follow-up at 6 weeks and 3, 6, 12 and 24 months post-op. The outcome measures chosen were the American Knee Society Scoring System (KSS) and three simple dichotomous yes/no questions concerning the ability to rise from a chair without the use of the arms and the pain experienced during this activity.

Despite ongoing attempts, the KSS has not been validated since its publication in 1989 (Insall et al, 1989). Pragmatically the scoring system is a useful clinical tool as it aggregates weighted scores for pain, range of motion, stability, alignment, and functional ability and has been widely used as an objective measure. Lingard et al (2001) however found the KSS to display poor face validity and differing levels of responsiveness between the clinical and functional components of the score, the former being responsive, the latter not. A further limitation is that the physical examination score is subject to misrepresentations, as poor correlation among the items of the KSS clinical score makes it possible for two very different patients to receive the same score. Lingard et al cite the example of, a knee score of 80 points being given to a patient who has no pain, 0° to 25° of knee flexion, normal

alignment, and no instability, or to a patient who has mild or occasional pain on walking and stair-climbing, 0° to 130° of knee flexion, normal alignment, and no instability. Clearly, these patients had very different results post surgery.

No difference in KSS was observed between groups in the Mahoney et al study. A larger proportion of the single radius group could rise form the chair unaided, and reported less pain at 6 weeks and 6 months post-op ($p<0.001$). At 1 year, 74% of the multi radius patients could complete the test unaided compared to 89% in the single radius group, which was reported as being significant at $p = 0.01$. At 2 years, the proportions were 73% and 90% respectively, with the significance having jumped dramatically to $p = 0.003$, based on only a 2% change in the number of patients that could achieve this function. There is no description of the quality of movement throughout the test, or the time taken to achieve the goal.

Despite the test consisting of an un-validated observation of an activity, Mahoney et al consider the chair rise test to be the more focused assessment of extensor mechanism function than their other outcome measure, as the KSS evaluates only walking and stair climbing and does not specify the location of pain experienced with these activities, thus conclusions of differing outcome between groups are drawn. The authors acknowledge that ideally allocation would be random and an independent assessor, blinded to implant type, should conduct the physical tests, but justify the comparability of their cohort as pre-operative demographic variables (gender, age and body mass) were not statistically different.

Hall et al (2008) reported no difference in extensor mechanism function in a study that benefited from a randomised trial design. Conclusions were based on equivalent KSS and ability to rise from a chair with out the use of the arms as per the Mahoney study. Prospective allocation of 50 multi radius implants (PFC sigma, Johnson and Johnson) were compared with 50 single radius designs (Scorpio, Stryker Orthopaedics) at the same assessment time points as Mahoney et al (2002). No between group differences were observed at any assessment.

As with the Mahoney et al study, this trial suffers from poor outcome measures, most emphasis being placed on the ability or not to stand from a chair without the use of arm assistance. Prospective randomisation of implants enhances the validity of the results, however the trial is powered on the results of the previous Mahoney et al study with an assumed 15% difference between groups in ability to raise from the chair, and a beta of 80% at an alpha of $p = 0.05$ to detect this.

The powering of this RCT is possibly inadequate, as it was based on the between group differences detected by the Mahoney et al study that suffered from numerous design issues as discussed. While the equivalent results of this study should not be ignored, it may be that the trial is not sensitive enough to test its hypothesis with the expected between group difference of 15% and accepted chance of a type II error of 20%.

Gomez-Barrena et al (2010) performed a retrospective case-controlled study with 60 patients comparing 30 single radius implants (Scorpio, Stryker Orthopaedics) against 30 multi radius devises (NexGen, Zimmer Inc). Outcome was assessed with KSS and number of physiotherapy sessions required for postoperative rehabilitation. Further analysis of muscle performance (isokinetic dynamometry) stability of the knee (dynamometric balance platform) and gait analysis was also reported.

The single radius group demonstrated better KSS and fewer physiotherapy sessions were required to reach post-operative goals. They concluded that there was better active knee extension in the single radius and an enhanced flexion / extension ratio on isokinetic testing. The authors report quadriceps recovery to be a key aspect in knee and posture stabilisation. 'Enough recovery' was described as occurring in both groups, but better extensor recovery seen in the single radius group. No differences were observed in the stability assessments conducted. Of the gait analysis, only the contralateral limb was found to perform differently between groups, which the authors suggest to be of importance and related to quadriceps recovery in the single radius group. Of note, no statistical correction was made for multiple testing.

Another interpretation of the data presented is that the contralateral leg differences reported do not reach an appropriate level of significance.

Again procedural ambiguities in the study design confound the outcome reported. This was a single retrospective assessment of unilateral TKA patients at, broadly, 11 months post-op for each group (that the ranges differ, 7-13 months in the single radius group and 9-14 months in the multi radius, is not addressed). Patient selection is stated to have been from the total 186 patients operated on in a 12 month period, and implant usage at the preference of the operating surgeon. Strict inclusion criteria results in only a third of these being eligible to take part, with convenient symmetry of 30 patients per group. The statistical power of the study is not discussed.

The physical patient benefits suggested to result from the use of a femoral component with a single radius of curvature have not been thoroughly investigated. Studies have either focused on highly technical assessments on small numbers of patients with a single assessment post-op design and carry little clinical relevance (Wang), or consist of larger more clinical studies with prospective design, but simplistic outcome assessments of which the validity is dubious (Mahoney et al, 2002; Hall et al, 2008). The power of these studies to detect differences is also questionable. Gomez-Barrena et al (2010) correctly comment that the theoretical advantages of the single radius design remain to be clinically proven.

2.5 Muscle Regeneration

As recently as 1990, in their specialist text book of muscle physiology, Jones and Round note that a widespread belief that skeletal muscle is not able to regenerate following injury is only now being acknowledged otherwise.

This intrinsic regenerative capacity was first demonstrated in 1964 by Studitsky who showed that a muscle that was removed, minced then returned to the original site was able to regenerate into a functional muscle. This capacity is attributable to the muscle satellite cell (Bischoff, 1975; Schultz, 1976; Snow, 1978; Hawke and Garry, 2001).

Muscle satellite cells

Initially described in the frog (Mauro, 1961; Katz, 1961), then subsequently in humans (Laguens, 1963), these cells are named according to their location anatomically, on the surface of the myofibre, between the plasmalemma and basal lamina (Figure 2.4). The muscle satellite cell is an undifferentiated myogenic precursor that is activated and proliferates following injury to repair the damaged muscle fibre (Mauro, 1961).

Mature skeletal muscle is a relatively stable tissue, thus the homeostatic demand on satellite cells is comparatively low and the majority are quiescent pending activation (Schultz et al, 1978; Hawke and Garry, 2001). These cells are primarily responsible for muscle growth and repair from the early post-natal period, throughout life (Zammit, 2008).

Figure 2.4 - Satellite cell anatomical location, occupying a sublaminar position in adult skeletal muscle. In the uninjured muscle fibre, the satellite cell is quiescent, however when injured, they increase their cytoplasmic content and develop cytoplasmic processes that allow for chemotaxis of the satellite cell along the myofiber to the site of damage. The satellite cells can be distinguished from the myonuclei as it lies beneath the basal lamina but above the plasmalemma.
Bar = 1 µm

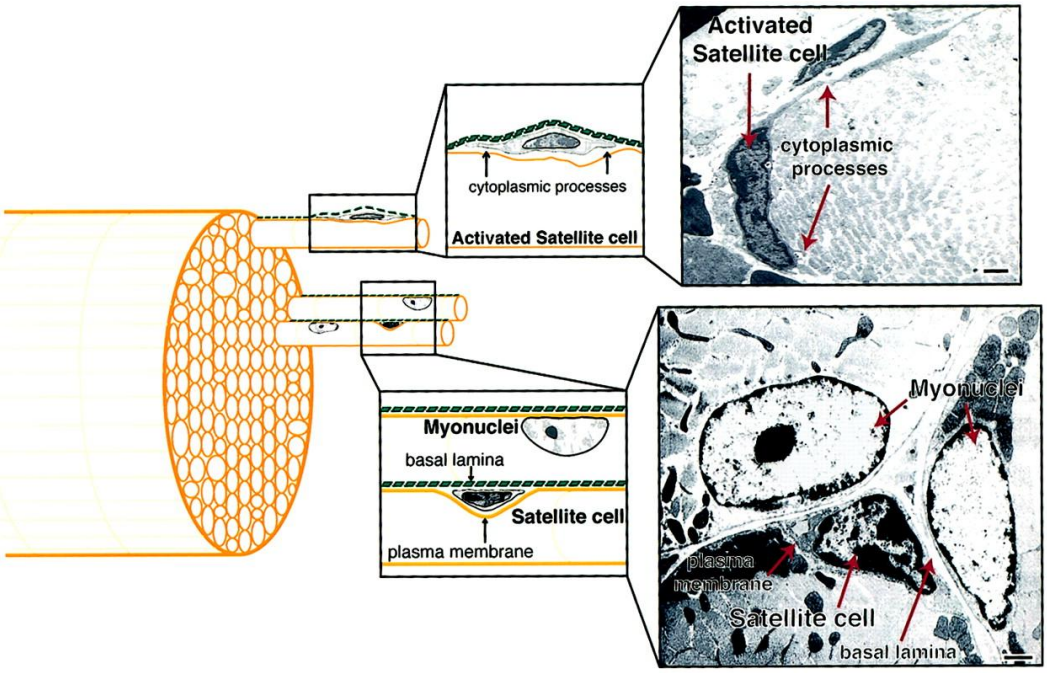


Diagram taken from Hawke T and Garry D. Myogenic satellite cells: physiology to molecular biology. *J Appl Physiol 2001;91-2:534-51.*

In normal muscle, satellite cells respond to regenerative cues such as injury or exercise, by proliferating to form myoblasts, which divide a limited number of times before terminally differentiating and fusing to form multinucleated myotubes (Morgan and Partridge, 2003; Wagers and Conboy, 2005) to provide new myonuclei for the homeostasis, hypertrophy or repair of the muscle fibres (Figure 2.5). These cells also self-renew in order to maintain a viable stem-cell pool that is able to respond to repeated demand (Hawke and Garry, 2001; Zammit et al, 2006).

Figure 2.5 - Satellite cell response to myotrauma. Upon activation some of the satellite cells will re-establish a quiescent satellite cell pool through a process of self-renewal. Others will migrate to the damaged area and, depending on severity of injury, fuse to the existing myofiber or align and fuse to produce a new myofiber. In the regenerated myofiber, the newly fused satellite cell nuclei is located centrally, but later migrate to assume a more peripheral location.

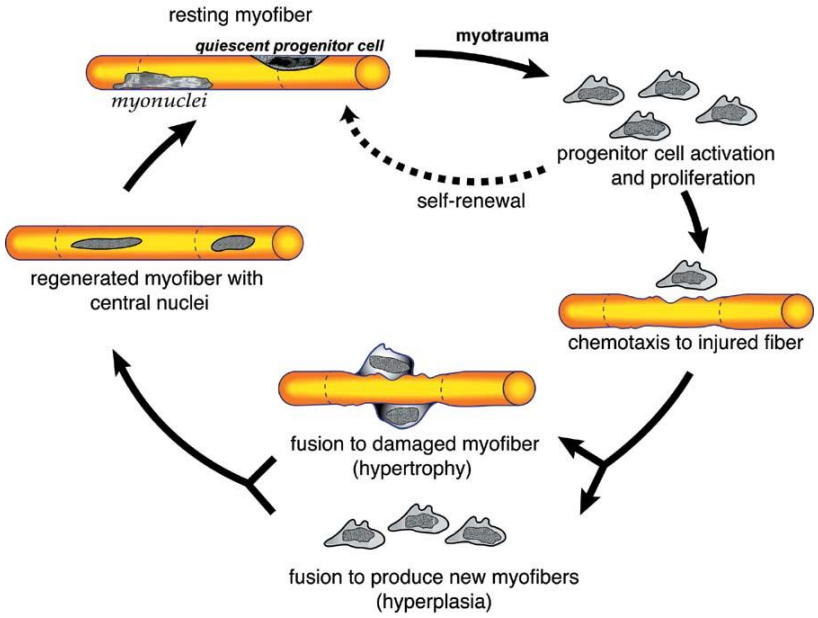


Diagram taken from Hawke T and Garry D. Myogenic satellite cells: physiology to molecular biology *J Appl Physiol* 2001 91:534-551.

Contribution to muscle hypertrophy

Boldrin et al (2010) in a review note that satellite cells become activated and increase in number in response to exercise. It has been demonstrated in the rodent model that satellite cells are required for hypertrophy of overloaded skeletal muscles (Snijders et al, 2009). Studies of 'unloading-induced' muscle atrophy, also demonstrate that satellite cells initially become activated (Ferreira et al. 2006), but eventually decrease in number (Hawke and Garry, 2001).

Human studies suggest that satellite cells of both young and old individuals respond similarly to exercise, increasing in number and activation status (Crameri et al, 2004;

Kadi et al, 2005; Verdijk et al, 2009) and contributing to muscle hypertrophy (Kadi et al. 1999; Kadi and Thornell 2000).

Regulation of satellite cell activity

The regulation of satellite cell proliferation and differentiation is not completely understood. There is general consensus that satellite cells are marked by the expression of the Pax7 gene and that this is primarily important in the cell activation process (Zammit, 2008; Buckingham, 2007; Seale, 2000). Activated cells express myogenic regulatory factors in a similar manner to muscle precursor cells during early muscle development (Morgan and Partridge, 2003). These myogenic transcription factors play a further role in the regulation of the differentiation process (Relaix, 2005). Current thought is that Pax7 regulates the entry of the satellite cells into the myogenic programme via the activation of the myogenic determination genes, MyoD and Myf5 (Buckingham, 2007).

Boldrin et al (2010) in a recent review suggest a working model of satellite cell activation and progression through the myogenic program. Quiescent satellite cells express Pax7 and Myf5. Upon activation, they up-regulate MyoD and divide to produce a pool of muscle precursor cells (MPC). Satellite cell progeny then follow one of two fates. They either down-regulate MyoD and self-renew to provide a new satellite cell, or alternatively, differentiate by down-regulating Pax7, Myf5, and MyoD and expressing MRF4 and myogenin, eventually fusing either to form new myofibres or to repair damaged myofibres (Figure 2.6).

Figure 2.6 - Model of satellite cell activation and progression through the myogenic program. Quiescent satellite cells express Pax7 and the myogenic regulatory factor (MRF) Myf5. Upon activation, they divide to produce a pool of muscle precursor cells. The cell progeny then follow one of two fates, self-renewal to give rise to another satellite cell, or differentiation to repair damaged myofibers.

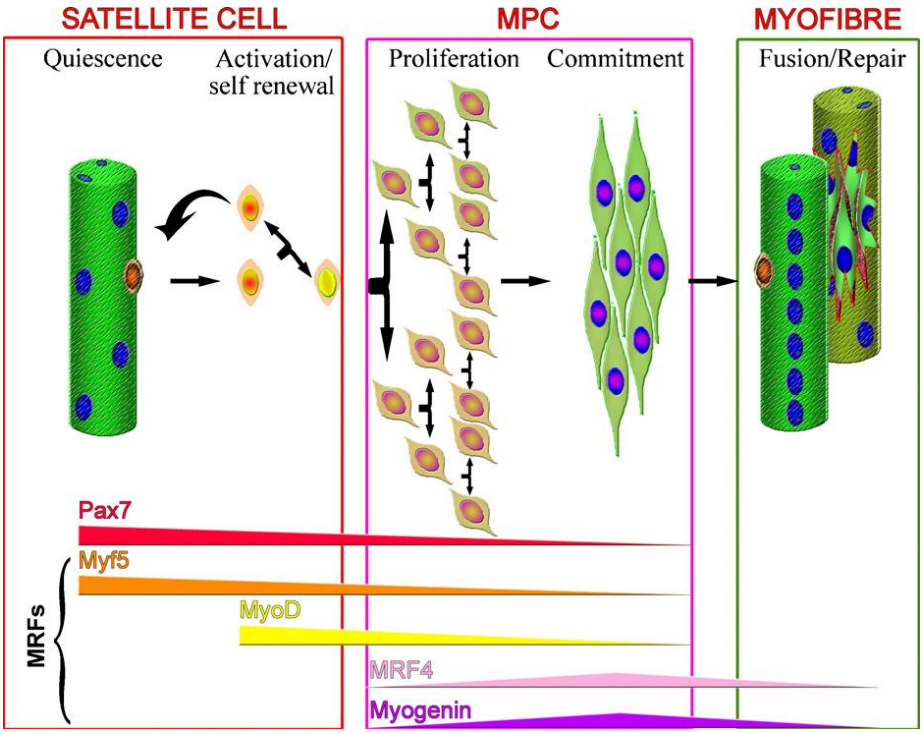


Diagram taken from Boldrin L, Muntoni F, and Morgan J. Are Human and Mouse Satellite Cells Really the Same? *J. Histochem. Cytochem.* 2010 Volume 58(11): 941–955

Potential therapies

There is much optimism that manipulation of satellite cells may result in therapies to treat muscular dystrophies and perhaps sarcopenia. Collins et al (2005) working in the mouse, transplanted single intact myofibres (containing approximately 7 satellite cells) into ablated muscle of immuno-compromised hosts. The grafted fibres were demonstrated to produce over 100 new myofibres, containing an estimated 25000 – 30000 differentiated myonuclei. These donor derived satellite cells persisted in the skeletal muscle for several weeks and could be reactivated and expanded in response to additional muscle injury.

It has been recently demonstrated that transplanting aged satellite cells into young hosts restores their regenerative capacity. Hall et al (2010) have shown (again in a murine model) that transplantation of myofibres with associated satellite cells, coupled with induced muscle injury, elicits lifelong enhancement in muscle mass, stem cell number and force generation in the host animal. When compared to the uninjured contralateral limb, a 50% increase in mass and a 170% increase in size was recorded that persisted for the lifetime of the mouse. Unfortunately, transplanted satellite cells have a limited capacity for migration, and are only able to regenerate muscle in the region of the delivery site (Relaix et al, 2005). As such systemic treatments or even the treatment of an entire muscle in this way is currently not possible.

Hall et al. (2010) speculate that an alternative therapeutic approach to enhance the regenerative capacity of the endogenous cell population may be possible by modifying the signalling pathways to mimic the environment of the transplanted satellite cells. While this idea is attractive, as the approach may provide therapeutic benefit in the absence of cell transplantation, there are many issues to resolve before any such therapies can be trialled. Hall et al. acknowledge that the underlying cellular mechanisms involved are not fully understood and that minimal work has been carried out with human tissue. Of further note, previous animal studies in stem cell and muscle transplantation have resulted in minimal physiological improvements (Cerletti et al, 2008).

Human studies

Much of the research into satellite cell activation and proliferation has been conducted in the murine model, with only a handful of authors conducting human studies. Boldrin et al (2010) comment that the lack of specific satellite cell markers in the human has led to equivocal, and sometimes contradictory, reports on their presence and number in human muscle sections.

In the few human studies that have been conducted, satellite cells have generally been identified with immunohistochemical antibody staining. Leu19, which recognizes CD56 (or neural cell adhesion molecule, NCAM) (Schubert et al, 1989), being the most common antigen. NCAM expression is not satellite cell specific (Schubert et al, 1989; Mechtersheimer et al, 1992), but has been extensively used for identification of satellite cells in human muscle (Schubert et al, 1989; Illa et al, 1992; Kadi et al, 1999; Kadi et al, 2006; Kadi and Thornell, 2000; Charifi et al, 2003; Mackey et al, 2007).

Although Pax7 is considered essential for the specification of satellite cells, it has been suggested that Pax7 staining does not account for all human satellite cells (Boldrin et al, 2010). This suggests that the use of Pax7 staining as a marker for satellite cells could, lead to an underestimation of total cell count. Previous authors however (Reimann et al, 2004; Verdijk et al, 2007) have demonstrated 96% of NCAM-positive satellite cells were also Pax7 positive, the remaining cells suggested as being either activated or differentiating, suggesting Pax7 to be a reliable marker. Authors now generally either use multiple antibody markers or a single marker in combination with the anatomical location to confirm the positive stain (Lindstrom and Thornell, 2009).

Human muscle regeneration and the aging process

A progressive loss of muscle mass and function, accompanies aging (Jones and Round, 1990; Kadi and Ponsot, 2010; Grounds, 1998), resulting in a decreased ability to generate power and force due to loss and change of contractile properties of the motor units in muscle (Lexell, 1995). Typical loss of muscle mass and strength between the ages of 25 and 80 has been estimated at between 33 and 50% (Lexell, 1993; Brooks and Faulkner, 1994). This loss of strength is directly associated with limitation in mobility and physical performance (Whipple et al, 1987). This in turn is associated with a high incidence of injury, and a loss of quality of life (Ryall et al,

2008). With increasing age, there is also a decline in the ability of skeletal muscle to regenerate (Kadi and Ponset, 2010; Grounds, 1998; Lexell, 1995).

As noted there is little experimental data on human satellite cells. The few studies that have been conducted though provide reasonable evidence that satellite cell content in skeletal muscle reduces with age (Renault et al, 2002; Kadi et al, 2004; Verdijk et al, 2007). Early work with electron microscopy (to determine the position of the satellite cell between basal lamina and sarcolemma of the muscle fibre) suggested no difference in satellite cell numbers in samples of vastus lateralis, reporting equivalent numbers in young and elderly men and women (Hikida et al, 1998; Roth et al, 2001). More recent investigations however, with immunohistochemistry and light microscopy suggest a reduction in cell number. Boldrin et al (2010) describe the use of electron microscopy for quantification as technically demanding and not suited to studying large portions of muscle, as the number of fibres that can be assessed is much lower than with immunohistochemistry. Electron microscopy does allow a detailed morphological analysis of the myonuclei and satellite cells, however the reported number of fibres assessed per individual is around 4 times less than with the immunohistochemistry method.

In contrast to the early investigations, Renault et al (2002) report fewer satellite cells in an older compared to young cohort. Biopsies were obtained from the masseter and biceps brachii muscles in 6 young (age range 20-28 years) and 6 old (58-83, years) subjects. Satellite cells were identified by an IHC staining protocol using NCAM and cell location. They further evaluated regenerative history of the individuals by measuring telomere length.

Telomeres are repeated DNA sequences located at the end of all chromosomes. They are small sections of non-coding DNA sequences, and play an important role in recombination function (Haber and Thorburn, 1984). During DNA replication, DNA polymerase is unable to copy the terminal segment of each DNA strand, which

results in shortening of the chromosomes at each round of cell division (Harley et al, 1990). In somatic cells, telomere length decreases regularly with cell division. In skeletal muscle, since nuclei are added to muscle fibres at various times during fibre regeneration, the mean telomere length reflects this heterogeneity, while the minimum length reflects the most recent nuclei incorporated, thus the length of telomeric DNA has been proposed as a good indicator of regenerative history (Renault et al, 2002). Using this approach Decary et al (1997) showed a small decrease in minimum telomere length (11 base pairs per year) in human quadriceps samples between 9 months and 86 years of age, whereas a comparatively large decrease (187 base pairs per year) was detected in children with muscular dystrophy (Decary et al, 2000).

Renault et al (2002) did not find any significant difference in telomere length in either the masseter or biceps between the old and young groups. The number of satellite cells reduced with age, while the number of myonuclei remained constant. They comment that these results suggest that a decrease in regenerative potential (as demonstrated by a decrease in the satellite cell number) is not accompanied by an excessive turnover of myonuclei, as the length of telomere fragments remained constant.

A decrease in the number of muscle fibres has been reported in the older population (Larson, 1978; Lexell 1988). This would suggest a loss of myonuclei and satellite cells, and would not explain the selective loss of satellite cells found in the study by Renault et al (2002). Renault et al thus speculates as to a deficit in the restoration of the 'pool' of satellite cells that return to quiescence following activation and division (Figure 2.5), and further that this decrease is probably related to continual low level turn over of satellite cells with aging and repair – which would progressively exhaust the proliferative capacity of some of these cells.

Kadi et al (2004) reported significant difference between young and old groups in a much larger cohort using immunohistochemistry staining for NCAM / CD56 and

position under the basal lamina. Biopsies of tibialis anterior were obtained from 15 young male and 16 young female (age range 20-32 years) and 13 old male and 14 female (age range 70-83 years) in subjects free from musculoskeletal complaints. The total number of myonuclei was found to increase with aging, however significantly reduced numbers of satellite cells per muscle fibre (approximately 40%) were observed in the older population, no gender effect was observed.

Verdijk et al (2007) found a specific satellite cell reduction in type II muscle fibres in the elderly. They obtained biopsies bilaterally from the vastus lateralis of 8 young (age 20 +/-1 years) and 8 elderly (age 76+/- 1 years) men, and assessed by an immunohistochemistry protocol for Pax7 and corroborated this count in a subset of tissue with a separate stain for NCAM / CD56. Previously Kadi et al (1995) had reported that no difference was found in satellite cell number between fibre types in the young subjects. The Verdijk study corroborated this finding, however noted a 44% lower content of satellite cells in type II fibres compared to the type I fibres in the elderly group. The satellite cell content in the type II fibres of the older group was also significantly lower when compared to the type II fibres of the young group. Verdijk et al suggest that this fibre type specific decline in satellite cell content might represent an important factor in the aetiology of sarcopenia.

Together these studies suggest that satellite cell number is reduced with aging, although it should be noted that the combined number of subjects in the 3 studies was only 45 young and 41 older individuals and that the biopsy site incorporated a variety of individual muscle groups.

Verdijk et al (2007) concede that the potential causes and clinical relevance of these changes in muscle satellite cell content with aging can only be speculated upon. It is also difficult to draw a direct causal link between satellite cell loss and age related muscle atrophy from this data. It is to be expected that progressive exhaustion of the satellite cell pool with advancing age should impair muscle regeneration; however mouse studies suggest that the inability of aged muscle to regenerate is due to an

inhibitory effect of the local environment on regenerative capacity (Conboy et al, 2005; Carlson and Faulkner, 1989). It is also possible that other factors, such as nervous system degeneration are relevant. This neurogenic process, associated with aging, has been reported to lead to a reduction in muscle mass (Dutta and Hadley, 1995). Interestingly, it has also been demonstrated that chronic denervation is associated with a reduction in local muscle satellite cell content (Rodrigues and Schmalbruch, 1995).

Another factor that may complicate the relationship between age and cell number is the effect of physical training. The activity levels and training state of individuals, irrespective of age, can affect satellite cell numbers (Chafri et al, 2003). As such, the cell numbers reported in the aging studies may reflect age related declines, or may be associated with activity levels that commonly reduce with advancing age.

Long term strength training has been shown to increase satellite cell numbers. In a study with high-level powerlifters, trapezius muscle satellite cell content was 70% higher than that of control subjects. Further studies investigating the short term response of satellite cells to exercise have demonstrated increases in both young (Kadi and Thornell, 2000) and elderly (Mackey et al, 2007; Chafri et al, 2003; Roth et al, 2001) populations. Chafri et al (2003) found a 29% increase in the number of satellite cells in elderly men in response to a cycle ergometer exercise program while Mackey et al (2007) reported an increase following a 12 week resistance exercise programme in elderly.

The regenerative response of skeletal muscle following exercise has been well reviewed by Grounds (1998) who described different levels of muscle damage that determine the cellular response. In ‘sub-lethal’ damage that was insufficient to provoke regeneration (examples of which were eccentric exercise with disruption of myofibrilar structure - particularly the z-bands), the response was macrophage activity and inflammatory response without associated myofibre necrosis and regeneration. In this situation, there was no need for satellite cell replication and

delivery of new myoblasts. The classic regeneration response of necrosis and regeneration follows more substantial damage, usually after intense or unaccustomed exercise. In this situation, the damaged myofibre is sealed off by new sarcolemma formation, and then an influx of inflammatory cells follows prior to satellite cell proliferation and fusion to repair the damaged segment.

Clinical relevance

While the influence of the muscle satellite cell has been considered with reference to microvascular free flap reconstruction for salvage of limbs following complex fractures and soft tissue trauma (Kauhanen et al, 2003), surprisingly, there has been no investigation of the role of the satellite cell in an orthopaedic population, and no consideration as to its influence in post-operative recovery. Specifically, any link between the regenerative potential of skeletal muscle, intrinsic satellite cell number and post-operative recovery has been neither investigated nor established in a knee arthroplasty population. This population is particularly relevant due to the substantial influence of the quadriceps on post-operative outcome (Greene and Schurman, 2008; Aspden, 2008).

Below normal muscular performance has been observed in both rheumatoid arthritis and osteoarthritis populations (Beals et al, 1985; Tiselius, 1969), and in particular specific weakness in the quadriceps group was noted (Nordesjo et al, 1983). The association of strength loss with osteoarthritis is traditionally thought to be that of disuse atrophy, secondary to joint pain. The muscle weakness can lead to joint instability which can contribute to mechanical overload of the joint (Aspden, 2008). This association has however now been questioned, and it has been suggested that muscle motor and sensory dysfunction may actually be a causative factor in knee osteoarthritis (Hurley, 1999).

Asmussen and Heeboll-Nielson (1962) and then Larson (1978) showed the age related decline in muscle strength to be most conspicuous in the proximal muscles of the lower limbs, while the musculature of the back and hand grip muscles were less affected. If one assumes the quadriceps and hamstrings to broadly consist of type II, phasic muscles and the hand grip and back muscles more of a mixed fibre type, this may suggest a relevance of the specific fibre type atrophy previously outlined. It may also though reflect a specific detraining of the lower limbs compared to the upper body musculature through decreased weight bearing activity. However this is difficult to ascertain clinically.

Barton and Morris (2003) in a review of strategies to counter muscle atrophy note two conceptual types of atrophy – acute and chronic. Acute atrophy is generally associated with disuse and chronic atrophy with aging and sarcopenia. While there are some similarities (reduced muscle mass and fibre size) there is also an important difference, in that fibre properties tend to shift toward faster fibres in acute (disuse) atrophy, whereas in chronic (ageing) atrophy a decrease of both in total fibre number and selective atrophy of the fastest, most powerful fibres is prevalent (Lexell, 1993). Applying this general classification to the knee arthroplasty population is difficult, as they are likely to demonstrate a mixed picture of both chronic and acute atrophy, though it reasonable that a reduction in both fibre number and specifically of type II fibres can be expected in many cases.

Those undergoing TKA are likely to have substantial associated muscle atrophy. How this affects recovery following knee replacement is unclear. Mizner et al (2005a) have demonstrated that the early/acute phase of recovery following knee arthroplasty may well be related to the voluntary activation of muscle. In this early post-operative phase both pain and muscle spasms complicate the picture. The time course of the inflammatory response and of muscle regeneration would suggest that any satellite cell influence may be related to longer term changes in function.

Following TKA, many patients are able to rehabilitate to enhance physical function and return to their desired level of activity, which may include gentle sports such as bowls, golf and doubles tennis (AAOS, 2010; BOA/BASK, 2010). Others find it difficult to regain post-operative function which is known to be linked to pre-operative quadriceps strength (Faulkner et al, 2010; Mizner et al, 2005b, Lingard et al, 2004; Lamb and Frost, 2003) though the mechanisms that modulate the relationship between muscle strength and post operative function are yet to be elucidated.

3 The Influence of Single Radius Femoral Component Design on Post-operative Function: The TRIMAX RCT

3.1 Introduction

TKA is an established and successful procedure (NIH position statement, 2003), however although the majority of patients do well, others report a variable outcome and moderate rates of dissatisfaction have been reported (Wright et al, 2004; Robertson et al, 2000). Jacobs et al consider that technical developments may potentially improve this procedure, which is already successful (Jacobs, et al, 2007a), but while many such developments in implant design have been aimed at improving quality of life, there is little consensus as to whether these design changes affect patient outcome (BOA/BASK position statement, 2010).

The single radius design is one such concept that has yet to be proven to be of benefit to patients. The Triathlon® prosthesis is a new implant developed with a single radius of curvature femoral component that is suggested to benefit patient recovery and subsequent function through enhanced reproduction of normal femoral kinematics and a longer moment arm for the extensor mechanism (resulting in a decreased requirement for muscle work to extend the knee).

Adequate extensor mechanism function is considered a prerequisite for participation in many activities of daily living and in promoting a positive clinical outcome (Noble et al, 2006; Silva et al, 2003). Quadriceps weakness is often present in patients receiving a TKA and has a substantial impact on movement patterns and performance during functional tasks (Mizner et al, 2005c). Previous studies (Gomez-Barrera et al, 2010; Hall et al, 2008; Wang et al, 2006; Mahoney et al, 2002) have assessed the use of single radius of curvature designs, though not with the Triathlon® prosthesis specifically and reported conflicting results as to patient benefits. As has been previously discussed (in Chapter 2), these studies all suffer from a range of

limitations including poor trial design and lack of statistical power to determine differences between the implant groups investigated.

Clear guidance is currently lacking as to the methodology required to determine post-operative outcome following TKA. As previously discussed (Chapter 2), the most thorough analysis will likely incorporate both patient reported outcome measures and direct functional testing. In this case direct functional testing is essential to specifically assess the extent of the influence of the enhanced moment arm of the extensor mechanism on subsequent muscle function.

The aims of this chapter were (1) to assess whether the difference in implant design influenced overall patient functional outcome, and (2) whether the design change in axis of rotation that lengthens the moment arm of the extensor mechanism specifically influenced the function of the extensor mechanism and the ability to generate muscle power.

To determine if any differences in patient outcome can be attributed to the differing implant design, a randomised controlled trial was designed comparing the single radius Triathlon® TKA against a traditional multi radius design, the Kinemax® TKA (both Stryker Orthopaedics, Mahwah, New Jersey). The Kinemax® implant is an example of the traditional condylar femoral component design (Pradhan et al, 2006) and has demonstrated a good track record of success (Pradhan et al, 2006; Wright et al, 2004; Back et al, 2001).

Due to the lack of clarity as to best appropriate post-operative outcome, a multi-modal trial design was established to reflect the various components of patient outcome. The primary outcome measure selected was the Oxford Knee Score as it has been developed specifically to measure the outcomes of knee replacements, is widely employed in orthopaedic research as a primary outcome measure, and has been shown to perform very well compared with alternative tools (KAT trial group,

2009; Garrat et al, 2004; Dunbar et al, 2000). A raft of secondary outcome measures encompassing pain and functional testing were also incorporated to the trial design to evaluate these aspects of patient outcome in a comprehensive manner (Beard et al, 2010; Stratford and Kennedy, 2006; Davies, 2002). Specific assessment of the power output of the extensor mechanism was included to the trial design as one of these additional outcome variables.

3.2 Methods

Study design and ethical approval

A prospective, double blind, randomised control trial was conducted to investigate the effect single radius of curvature femoral component implant design on subsequent patient outcome following knee replacement surgery.

The trial was powered based on the ability to show a difference of 3 points on the Oxford Knee Score, the value currently accepted to reflect a clinically meaningful difference (Murray et al. 2007). The minimally clinically important difference (MCID) is the smallest change in score which patients perceive as meaningful, and which would cause clinicians to change management (Fayers and Machin, 2000). MCID values have not yet been calculated for the Oxford Knee Score, however Murray et al (2007) comment that pending this, approximation can be derived by calculating half the standard deviation (SD) of change in score. The SD of change in OKS is typically reported as between 6 and 10 points, and thus the MCID is likely to lie between 3 and 5 points. Some studies however have reported lower scores, and thus the real difference may actually be less than this. Clinically, the 3 point difference is approximated as the meaningful value and is widely used to power clinical studies. A power calculation was conducted with an alpha of 0.05 and beta of 0.8, which resulted in 102 patients per arm required to account for a total drop out of 10% at 1 year.

Local ethical approval was granted for the study by Lothian Research Ethics Committee and NHS Lothian Research and Development Management.

Patient recruitment

Patients were recruited from the planned operating lists of six consultant orthopaedic surgeons in the arthroplasty service at the Royal Infirmary of Edinburgh. All six have substantial experience of using both implants, and were content to use either for the duration of the trial.

Suitable patients were approached at the time of pre-operative assessment, approximately 2 weeks prior to surgery, and were recruited to the trial through informed consent. Once recruited, the lead surgeon allocated the patient either implant using a computer randomisation program. Both the patient and researcher were blinded to the allocation, and remained so throughout the trial period. A colour code was used to represent the respective implant types, and was used to instruct the subsequent surgical teams as to which implant trays should be prepared for the procedure.

A pragmatic trial design was adopted that best reflects the wider clinical population, and specific recruitment criteria applied. Inclusion criteria for participation was that patients were suffering from osteoarthritis of a severity that was suitable for joint arthroplasty, but were otherwise healthy (i.e. not suffering form any co-morbidities that would affect their post operative recovery or subsequent performance, such as cardiovascular or neurological disease).

Previous joint replacement (in other major joints) was accepted, as was osteoarthritis in the contralateral knee or either hip, provided that it was not of a severity that would be considered for surgery and / or did not impact the patient's functional performance. Additional criteria was that the planned procedure was 'routine'

primary total knee arthroplasty, i.e. the first surgical procedure on the osteoarthritic joint, performed with standard versions of the trial implants, without the addition of any augments to address compromised bone stock. Finally patients had to be physically and mentally able to comply with the post-operative testing and trial protocol.

Operative technique

Both implants were inserted utilising the same technique and standard instrumentation for each system; using intra-medullary referencing for the femur and extra-medullary jigging for the tibia. Patello-femoral resurfacing as part of TKA is not routinely practiced at the arthroplasty unit at the Royal Infirmary of Edinburgh, and no patients recruited to this study underwent patello-femoral resurfacing. Components were cemented in all cases. Implant alignment was specifically set with femoral rotation of 3 deg of external rotation using posterior referencing and rotation confirmed by assessing the transepicondylar axis and Whitesides line.

Follow-up and patient withdrawal

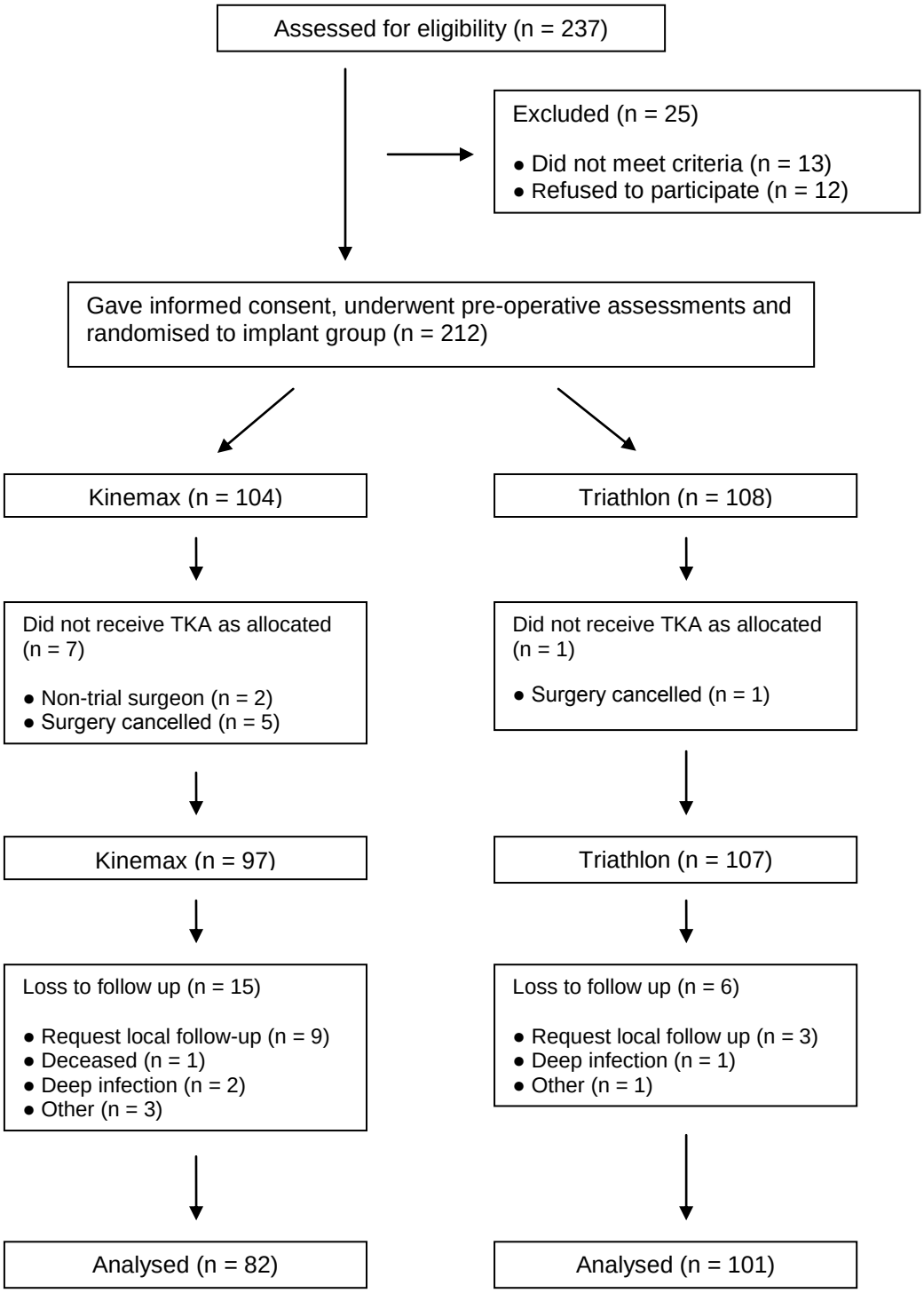
212 patients were recruited to the study, although 24 were subsequently withdrawn and a further five patients experienced complications that prevented them completing the 12 month assessment protocol. As a result 183 patients completed the full study protocol (reference trial flow chart).

Of the withdrawn patients, six procedures were cancelled or delayed beyond the trial recruitment period due to pre-operative medical complications. A further two procedures were transferred to ‘non-trial’ consultants as a part of a waiting list initiative. Due to their specific practice, these surgeons were not happy to use the trial implants, and opted instead to use another. One patient died during the follow-up period (unrelated to the surgery), another patient acquired a deep infection on the

ward post-operatively that led immediately to a 2-stage revision procedure, one patient experienced a peroneal nerve complication during surgery that prevented functional testing post-op, and a further patient subsequently revealed that she suffered from chronic pains in multiple joints. This was not divulged at the recruitment stage, and would have excluded her from participating in the trial. Despite previously consenting to the trial protocol with full information of the post-operative procedures, 12 patients requested post-operatively that they be followed up by clinical teams at their local satellite hospital which meant that they did not attend attending the Royal Infirmary orthopaedic research facility; these patients were withdrawn.

Of the 5 later complications, two patients were diagnosed with deep infections and had revision procedures. One had a post-operative flare-up of previously undiagnosed rheumatoid arthritis that required management with disease modifying drug therapy. One developed a progressive foot neuropathy that prevented functional assessment, and one could not be contacted for final follow-up at 12 months.

Trial participation flow chart



Outcome assessments

Patients were assessed pre-operatively, then post-operatively at routine outpatient clinical reviews at 6 weeks, 26 weeks and 52 weeks following surgery in a local clinical testing facility at the Royal Infirmary of Edinburgh. A comprehensive physical outcome testing protocol was used to evaluate patient function. This comprised the Oxford Knee Score, range of motion, pain score, timed test of activities of daily living and lower limb power output. All tests were carried out by the same researcher in the same manner.

Oxford Knee Score

The importance of a joint-specific tool lies in the ability to isolate the function of a single joint from the overall functional picture. The Oxford Scores are well validated as sensitive tools that are accepted by patients' and surgeons to gauge pain and functional outcomes (Dawson et al, 1998; Murray et al, 2007; Garrat et al, 2004). The Oxford Knee Score (OKS) has undergone one of the most thorough assessments of reliability and validity of all the tools used to assess outcome following TKA (Davies et al, 2002) and it use recommended (Davies et al, 2002; Dunbar et al, 2001).

The OKS consists of 12 equally weighted questions addressing pain and functional activity, each scoring form 1 – 5. Best possible scores are 12 and worst 60. Specific attempt was made during the construction of the OKS to minimise the influence of comorbidities (Murray et al, 2007).

As relative change in OKS is considered a more valid assessment of outcome than the absolute OKS (Murray et al, 2007; Price et al, 2010), the change in score value is the criteria of interest.

Range of motion

Active measures of flexion and extension were determined at all assessments by the same assessor using universal goniometry. A high level of accuracy has been previously demonstrated assessing knee range of motion with this instrument in the clinical setting (Watkins et al, 1991) and specifically in patients following TKA (Jakobsen et al, 2010).

The twelve-inch universal goniometer was chosen due to high intratester reliability (Brosseau et al, 1997). Brosseau et al (2001) noted high levels of intra-rater reliability for repeated goniometric measurements of knee range of motion in patients with restricted movement (ICC for flexion 0.997, and for extension 0.975) when evaluated against gold standard radiographs. Inter-tester reliability was found to be less accurate, and it has been recommended that the same assessor conduct all assessments when assessing active range in such patient groups.

Functional assessment

The aggregated locomotor function (ALF) score is a simple measure of observed locomotor function using timed tests of walking, stairs ascent/decent and chair transfers. It has been previously demonstrated to be valid, reliable and responsive, and proposed as a tool with which to assess physical function and to quantify treatment response (McCarthy and Oldham, 2004). These three measures have extensively been used individually as outcome assessments, but it has been suggested that a composite aggregated time provides a better objective assessment of the patients overall functional capabilities (Hurley and Scott, 1998).

The test was conducted in the manner suggested by McCarthy and Oldham (2004). Specifically, patients were asked to walk over a flat eight metre course, ascend and

then descend a rehabilitation style platform consisting of seven fixed steps (four of 15cm and three of 20cm) using banisters if required and additionally walk two metres to a chair (with seat height of 0.46m), sit down, then immediately stand up and return to the start.

In all tests patients were instructed to move at their own preferred 'comfortable' pace. A mean of three repetitions of the walking and transferring assessments and four repetitions of the stairs assessment were recorded. The four attempts of the stairs assessment were to balance the differing step heights. By traversing the steps in both directions, the patients ascend and descend the different height steps twice (McCarthy and Oldham, 2004). Time was recorded in seconds using a handheld stopwatch (Zeon, UK); the same assessor conducted each assessment, preventing intra-tester reliability issues around the use of the stopwatch.

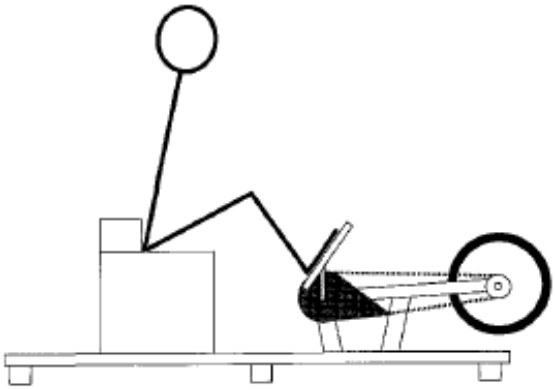
Lower limb power output

Specific assessment of the recovery of muscle power was assessed by use of a Leg Extensor Power Rig (Queens Medical Centre, Nottingham, NG7 2UH) which has been well validated for use with this population group (Lamb and Frost 2003; Robertson et al, 1998; Bassey et al, 1992). The Leg Extensor Power (LEP) Rig consisted of a seat and footplate connected through a lever and chain to a flywheel (Figure 3.1). Application of force accelerated the flywheel from rest, and output was recorded as maximal wattage (W) generated.

The test procedure followed that suggested by Robertson et al (1998) where the patient was seated in an upright position with arms folded, so as not to contribute to power generation. The distance between the pedal and seat was dependant on the length of the individual's lower limb. Seat position for the test was determined by positioning the patient in comfortable extension of the knee in conjunction with full depression of the foot pedal. The subject was instructed to move the pedal by

pushing the leg into extension with maximal effort. A command of 'push as hard and as fast as you can' was given prior to each effort as per the manufacturers instructions. Where continued progressive improvement in power output was observed, a maximum of 10 attempts was conducted. A minimum rest period of 20 seconds between attempts was enforced. Those not able to complete the test were assigned a score of zero as advocated by Lamb and Frost (2003). After each attempt, the force produced was recorded. The highest recorded output was used for analysis as suggested by Barker and Simpson (2004).

Figure 3.1 - Diagrammatic representation of the patient seated on the leg extensor power rig, demonstrating the leg position at the start of the push.



Adapted from: Bassey. E, Short A. A new method for measuring power output in a single leg extension: feasibility, reliability and validity 1990, *Eur j Appl Physiol* 60: 385-390

Power output was also recorded for the patient's contralateral leg to establish approximate 'normal' levels of output for that individual. Values were recorded at the 12 month assessment period to aid consistency and to standardise any influence of training effects through the rehabilitation process.

Pain assessment

Global knee pain severity (not activity specific pain) was assessed using an 11 point (0-10) numerical rating scale (NRS), where 0 represents no pain and 10 the worst

possible pain. The validity (Jensen and Karolyn 2001; Van Koff et al, 2000) and sensitivity (Williamson and Hogarth, 2005) of the NRS has been well documented, it is easy to administer and score and can also be used with a greater variety of subjects than other pain scales (Bellamy, 1993; Van Koff et al, 2000).

It has been suggested that using multiple measurements of pain status as opposed to a single value of 'current pain' may provide more realistic and meaningful measurements of pain intensity (Jensen et al, 1996). Separate assessments were made of 'worst pain' and 'perceived mean daily pain' as has been specifically recommended for use in OA clinical trials (Perrot et al, 2010). Patients were asked to state the severity of the worst and the average pain they experienced over the previous week. The validity of pain recall has been previously demonstrated over this timeframe, specifically with patients with osteoarthritis of the knee (Perrot et al, 2010).

Statistical analysis

Data was collated and analysed using SPSS version 17. Microsoft Office Excel 2003 and Minitab release 15 were further utilized to present data for superior graphical quality. All available data was included in the analysis to limit any effect of loss to follow up (Murray et al 1997).

The RCT was powered to determine between group differences in the OKS at 12 months, thus all outcome variables were assessed initially with Two-Way Repeated Measures ANOVAs to investigate the overall change between pre-op and twelve months assessment, and also the effect of implant group on the overall change, as was accommodated by the trial design. Statistical significance was accepted at $p = 0.05$.

Specific differences in outcome between the groups were assessed separately at the 4 assessment time points to investigate the changing outcome in the year following surgery. Independent samples t-tests or Mann-Whitney U tests (depending on the underlying distribution of the data) were used to assess between group differences at the four assessment points. Differences between the four assessments were assessed with Paired Samples t-tests or Wilcoxon Signed Ranks tests for repeated measures, depending on the normality of the underlying data. Multiple testing may lead to a type 1 error, thus the Bonferroni correction was applied to the level of accepted significance of $p = 0.05$ resulting in an accepted value of $p = 0.0125$ in each situation.

3.3 Results

Despite random allocation to ensure equal distribution of implant type, the withdrawn patients were unevenly split between groups, resulting in 101 Triathlon implants and 82 Kinemax implants being available for analysis (see trial flow chart).

The male to female ratio was equivalent between groups. The Triathlon group contained 37 male (36.6%) and 64 female (63.4%) patients, the Kinemax group 34 male (41.5%) and 48 female (58.5%) patients (Figure 3.2). The mean age of the trial cohort was 68.4 years, and was consistent between groups (Table 3.1).

Figure 3.2 - Bar chart demonstrating the trial split of patient gender by implant group

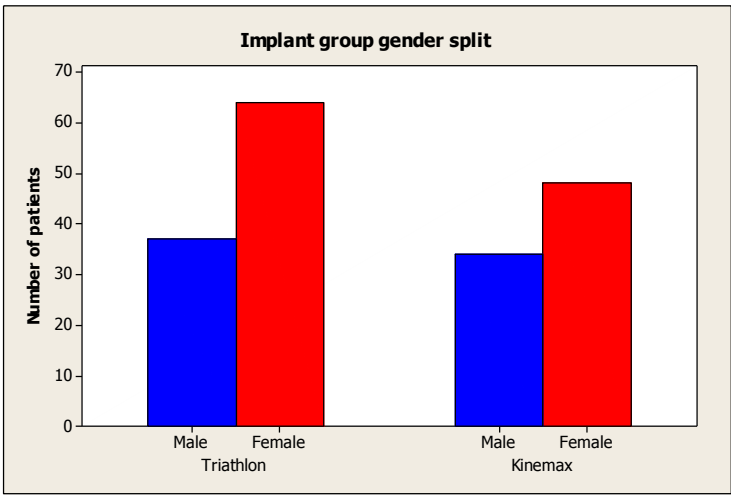


Table 3.1 – Mean age of patients at time of surgery by implant group.

Implant group	mean (+/-SD)	Range
Triathlon	68.9 (9.11)	46 – 92
Kinemax	67.7 (8.92)	33 – 84

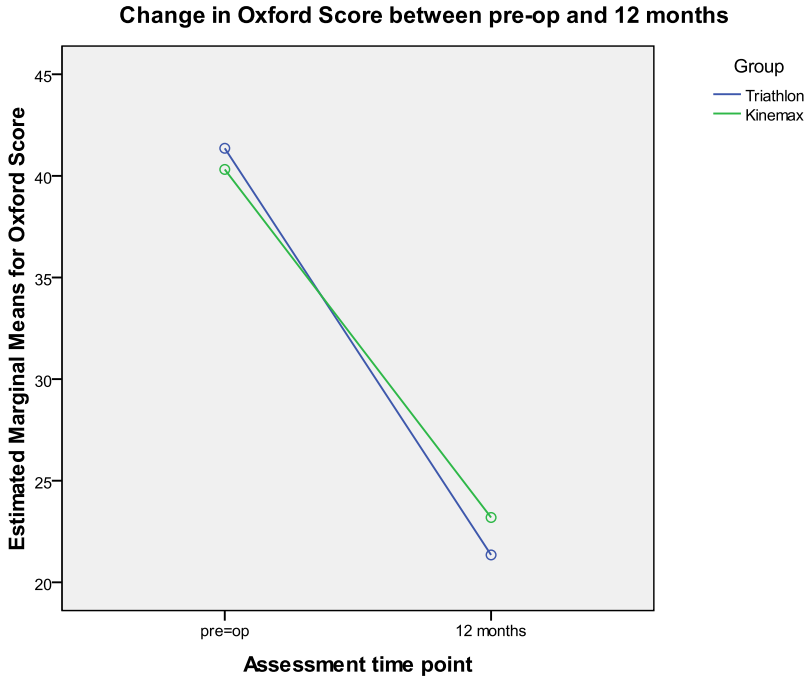
Oxford Knee Score

RCT results: 1 year outcome

The treatment effect of the TKA caused a left shift of the data post operatively, resulting in a non-Gaussian distribution when assessed at individual time points post-operatively. The repeated measures ANOVA however considered the change in score between the assessments, and this 'change data' was normally distributed. The mean change in OKS (difference between pre-op and 12 month assessment periods) was 20.01 (+/- 8.99) degrees for the Triathlon group, and 17.13 (+/- 9.35) for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess overall change in outcome between pre-operative and 12 month assessment periods and the influence of implant groups. There was a significant main effect of the TKA surgery reflected in a pre-operative to 12 month change in the Oxford Knee Score F (1, 172) = 709.5 p = <0.001. Further there was a significant interaction of the implant used and the change in Oxford Knee Score F (1, 172) = 4.28 p = 0.04 (Figure 3.3 and Appendix C for statistical output).

Figure 3.3 - Graphical illustration of the repeated measures ANOVA for OKS. The overall change and trajectory of change in both groups is highlighted. Although the overall change is similar, the triathlon group demonstrates greater overall change in Oxford Knee Score. See Figure 3.4 for ranges at individual time points.



Longitudinal change across the 4 time points

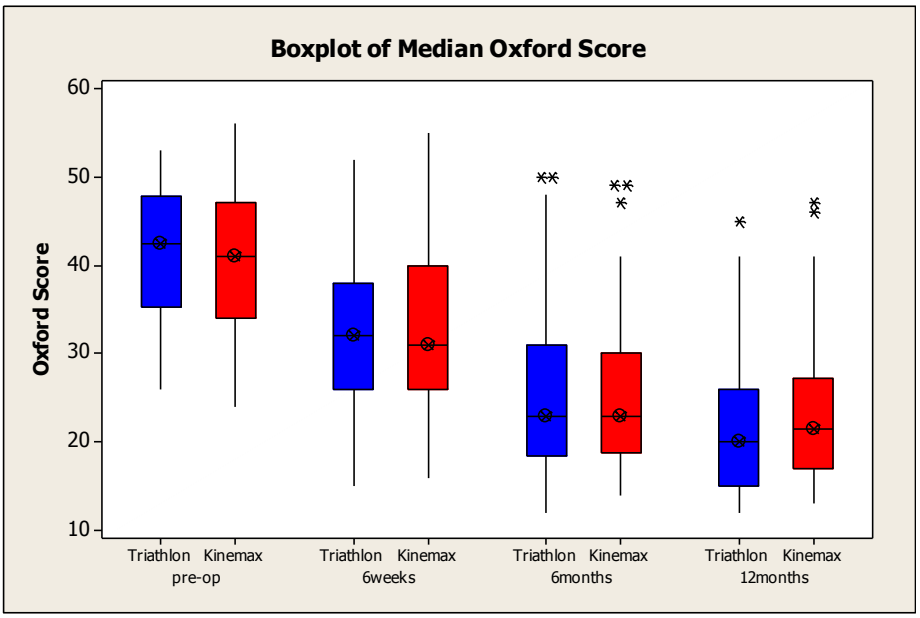
Specific analysis of the four assessment points highlighted the similar overall pattern of reduction in Oxford Knee Score (Table 3.2 and Figure 3.4). The change in score at each time point was similar, though the divergence was apparent at the 12 month assessment.

Table 3.2 – Oxford Knee Score

	Triathlon	Kinemax
Pre-op	42.5 (35.25, 47.75)	41.0 (34, 47)
6 weeks	32.0 (26, 38)	31.0 (26, 40)
6 months	23.0 (18.5, 31)	23.0 (18.75, 30)
12 months	20.0 (15, 26)	21.5 (17, 27.5)

Data presented as medians (+/- IQR)

Figure 3.4 - Boxplot of Oxford Knee Score, highlighting the similar pattern of reduction of OKS across the assessment period.



X denotes the median result and * an outlying data point.

Wilcoxon Signed Rank tests demonstrated statistically significant differences in Oxford Score between all assessment periods for both implant groups, p = <0.001 (Appendix C for statistical output), highlighting a step wise improvement in score over the first year post-operatively.

Mann-Whitney U-tests of between group differences at the four assessment periods highlight that while the overall change in score was significantly different between

groups, at no specific assessment point were the implant groups significantly different to each other (Table 3.3 and appendix C for statistical output). This again highlighted the similar pattern of improvement in patient reported score between groups.

Table 3.3 – Between group differences at specific assessments

	Z statistic	Significance
Pre-op	-0.80	p = 0.42
6 weeks	-0.17	p = 0.86
6 months	-0.12	p = 0.90
12 months	-2.18	p = 0.03

Significance accepted at p= 0.0125 when Bonferroni correction applied. No results reach required level of significance.

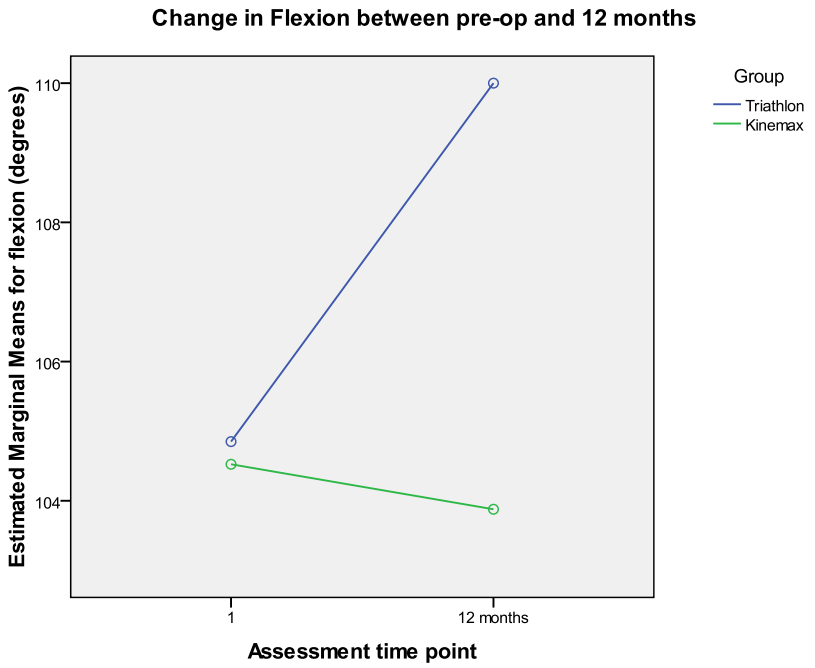
Range of motion – Flexion

RCT results: 1 year outcome

The data was normally distributed at all time points. The mean change in flexion (difference between pre-op and 12 month assessment periods) was 5.15 (+/- 12.37) degrees for the Triathlon group, and 0.65 (+/- 14.57) degrees for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKR surgery on the recorded flexion, $F(1, 180) = 5.08$, $p = 0.025$. There was also a significant interaction of the implant group and the change in flexion, $F(1, 180) = 8.418$, $p = 0.004$ (Figure 3.5 and Appendix C for statistical output).

Figure 3.5 - Graphical illustration of the repeated measures ANOVA for flexion, highlighting the differing trajectory of change in the implant groups. The Triathlon group achieves improvement in flexion following surgery, however this was only of the order of 5 degrees, while the Kinemax group broadly achieves pre-operative levels of flexion. See Figure 3.6 for errors at individual time points.



Longitudinal change across the 4 time points

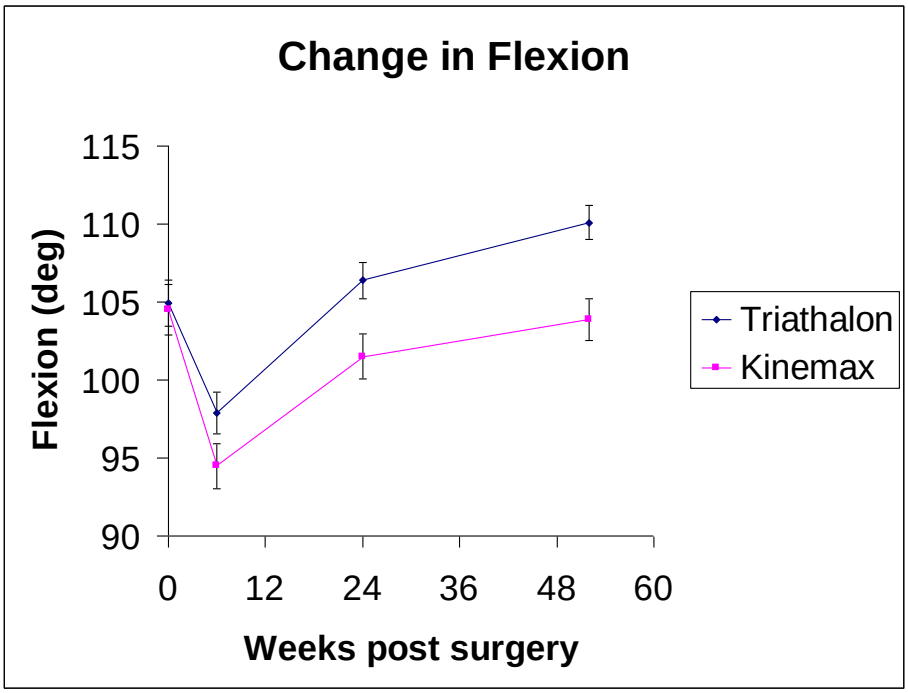
Specific analysis of the change at the four different assessments highlighted a similar overall pattern of change in flexion (Table 3.4 and Figure 3.6). Both groups presented with a mean pre-operative flexion of around 105 degrees, which reduced at first post-operative assessment, then recovered at later assessments. The differing amount of recovery of flexion between implant groups was also apparent, the Triathlon group achieving a flexion greater than pre-operative levels at 6 months, which subsequently improved further, the Kinemax group never regaining the pre-operative levels of flexion.

Table 3.4 – Flexion at assessment point.

	Triathlon	Kinemax
Pre-op	104.9 (14.8)	104.5 (14.5)
6 weeks	97.9 (13.3)	94.5 (13.0)
6 months	106.4(11.8)	101.5 (12.6)
12 months	110.1 (10.8)	103.9 (12.0)

Data presented as degrees of flexion mean +/- (SD)

Figure 3.6 - Change in flexion over the assessment period. Graph demonstrates the striking similarity of the change in flexion in both groups, and the superior results achieved by the Triathlon group.



Data plotted at assessment time points highlighting the proportional change in flexion over the year. Error bars represent the SEM.

Paired samples t-tests demonstrated statistically significant differences in flexion between all assessment periods for both implant groups, p = <0.001 (Appendix C for statistical output).

Independent samples t-tests of between group differences at the four assessment periods highlighted statistically significant differences between implant groups at 6 months and 12 months, the Triathlon group achieving greater flexion than the Kinemax group (Table 3.5 and also Appendix C for statistical output).

Table 3.5 – Flexion difference between groups at assessments

	Mean difference (95% CI)	Significance
Pre-op	0.33 (-3.98, 4.63)	p = 0.88
6 weeks	3.88 (-0.53, 7.29)	p = 0.09
6 months	4.87 (1.22, 8.52)	p = 0.009
12 months	6.19 (2.84, 9.57)	p = <0.001

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result.

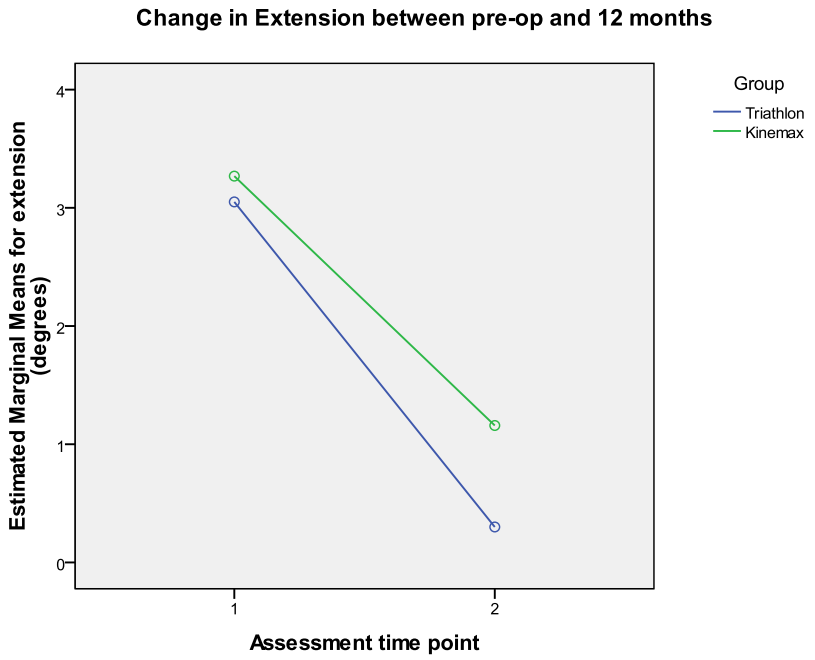
Range of motion – Extension

RCT results: 1 year outcome

The data was normally distributed at all time points. The mean change in extension (difference between pre-op and 12 month assessment periods) was -2.75 (+/- 5.09) degrees for the Triathlon group, and -2.11 (+/- 5.42) degrees for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKR surgery on the recorded extension, F (1, 180) = 38.7, p = <0.001. There was no significant interaction between the implant group and the change in extension, F (1, 180) = 0.67, p = 0.414 (Figure 3.7 and Appendix C for statistical output).

Figure 3.7 - Graphical illustration of the repeated measures ANOVA for extension. The reduction in mean flexion and similar trajectory of change in both groups is highlighted. See Figure 3.8 for errors at individual time points.



Longitudinal change across the 4 time points

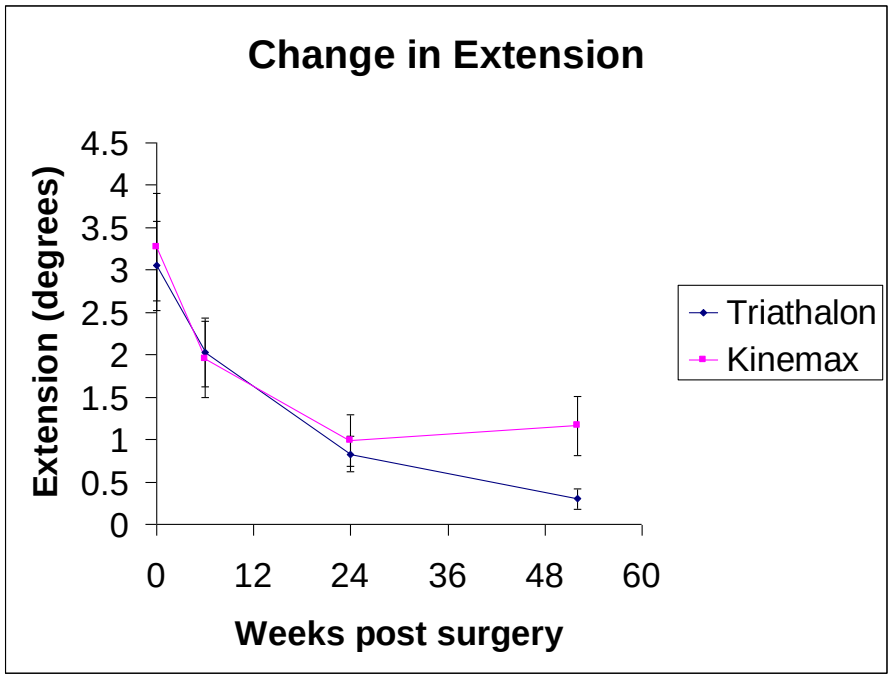
Specific analysis of the change at the four different assessment points highlighted the similar overall pattern of change in extension; both groups record a small mean limitation in extension that gradually reduced at subsequent assessments (Table 3.6 and Figure 3.8).

Table 3.6 – Extension at assessment point

	Triathlon	Kinemax
Pre-op	3.05 (5.24)	3.27 (5.70)
6 weeks	2.03 (3.97)	1.95 (4.03)
6 months	0.83 (2.07)	0.99 (2.61)
12 months	0.30 (1.21)	1.16 (3.32)

Data presented as degrees of extension from neutral mean (+/- SD)

Figure 3.8 - Change in extension over the assessment period. Graph demonstrates the similar early change in extension in both groups to 6 months, then the further improvement in the Triathlon group



Data plotted at assessment time points highlighting the proportional change in extension over the year. Error bars represent the SEM.

Paired samples t-tests demonstrate significant change in extension between 6 weeks and 6 months for both implants. The Triathlon group demonstrated further significant reduction between 6 and 12 months, whereas the Kinemax group extension plateaus (Table 3.7, Figure 3.8 and also Appendix C for statistical output).

Table 3.7 – Extension change between assessments

	Implant	Mean (95% CI)	Significance
Pre-op – 6 weeks	Triathlon	1.13 (-0.05, 2.30)	P = 0.06
	Kinemax	1.29 (0.09, 2.48)	p = 0.035
6 weeks – 6 months	Triathlon	1.22 (-0.06, 1.82)	p = <0.001
	Kinemax	1.09 (0.40, 1.78)	p = 0.002
6 months – 12 months	Triathlon	0.54 (0.23, 0.84)	p = 0.001
	Kinemax	0.09 (-0.27, 0.45)	p = 0.062

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicated a significant result.

Independent samples t-tests of between group differences at the four assessment periods highlighted that no differences exist pre-op, at 6 week or 6 month assessments, however the Triathlon group demonstrated a significantly improved mean extension compared to the Kinemax group at 12 months (Table 3.8, Figure 3.8 and also Appendix C for statistical output).

Table 3.8 – Extension difference between groups at assessments

	Mean difference (95% CI)	Significance
Pre-op	-0.22 (-184, 1.40)	p = 0.79
6 weeks	0.08 (-1.11, 1.28)	p = 0.894
6 months	-0.16 (-0.87, 0.56)	p = 0.667
12 months	-0.22 (-0.63, -0.10)	p = 0.017

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result.

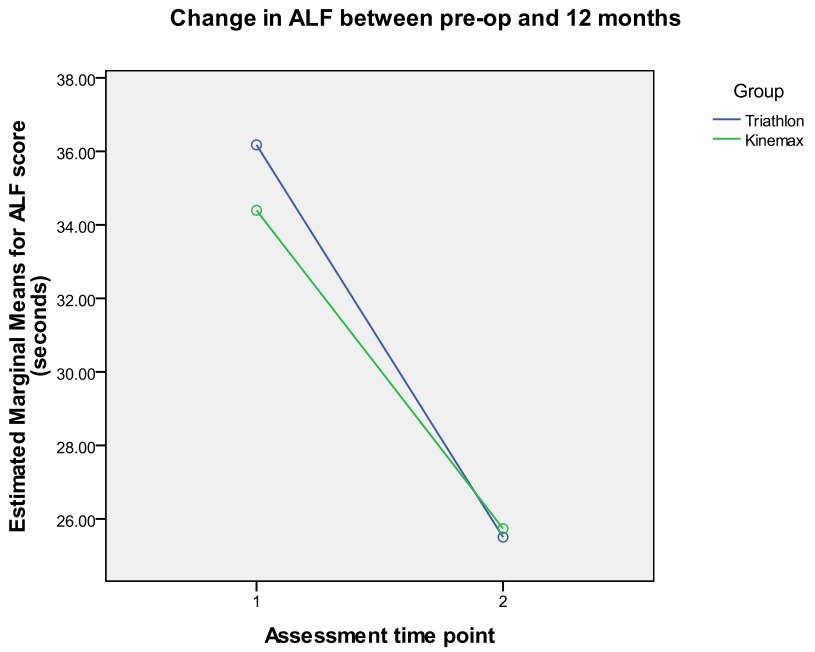
Timed functional assessment

RCT results: 1 year outcome

The data was normally distributed at all time points. The mean change in timed functional assessment score (difference between pre-op and 12 month assessment periods) was -10.67 (+/- 12.01) seconds for the Triathlon group, and -8.67 (+/- 10.93) seconds for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKR surgery on the timed functional assessment, $F(1, 181) = 127.0$, $p = <0.001$. There was no significant interaction of the implant group and the timed assessments, $F(1, 180) = 1.37$, $p = 0.243$ (Figure 3.9 and also Appendix C for statistical output).

Figure 3.9 - Graphical illustration of the repeated measures ANOVA for timed functional assessment. The overall reduction following surgery and the similar trajectory of change in both groups is highlighted. See Figure 3.10 for errors at individual time points.



Longitudinal change across the 4 time points

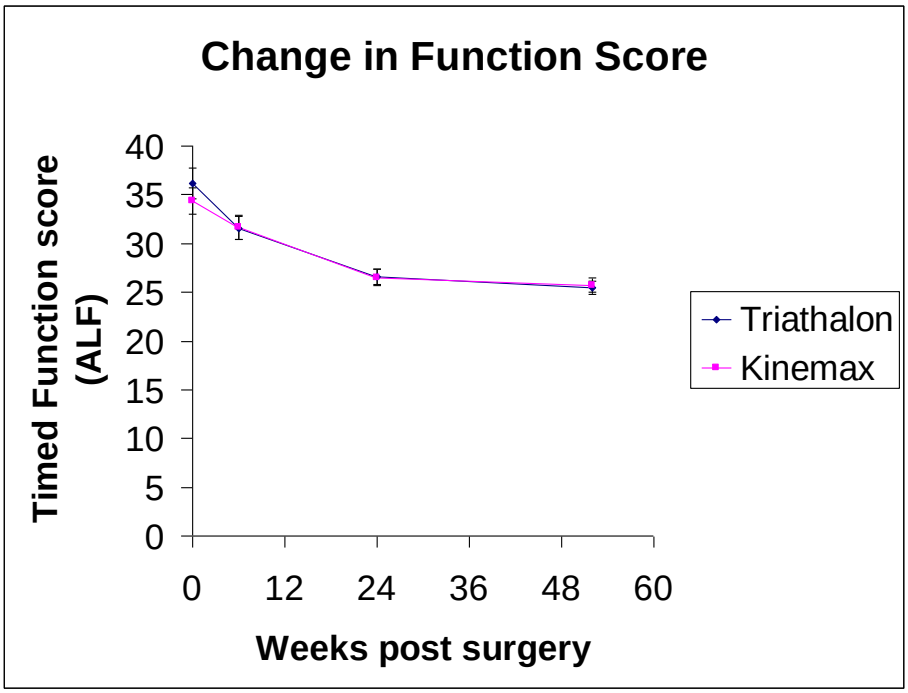
Specific analysis of the change at the four different assessment points highlighted the similar overall pattern of change in timed assessment. Both groups demonstrated a similar improvement in combined functional assessment in the year post surgery (Table 3.9 and Figure 3.10).

Table 3.9 – Timed function score at assessment point

	Triathlon	Kinemax
Pre-op	36.18 (3.06)	34.4 (2.67)
6 weeks	31.59 (2.37)	31.7 (2.41)
6 months	26.6 (1.55)	26.49 (1.67)
12 months	25.5 (1.33)	25.73 (1.37)

Data in seconds, presented as mean (+/- SD)

Figure 3.10 - Graph demonstrates the similar change in functional score in both groups. Data plotted at assessment time points highlighting the proportional change in extension over the year.



Error bars represent the SEM.

Paired samples t-tests demonstrated significant changes in timed functional assessment between all time points for the Triathlon group. Significant improvements in the Kinemax group were only observed between 6 weeks and 6 months (Table 3.10, Figure 3.10 and also Appendix C for statistical output).

Table 3.10 – Timed function change between assessments

	Implant	Mean (95% CI)	Significance
Pre-op – 6 weeks	Triathlon	4.64 (2.27, 7.01)	p = <0.001
	Kinemax	2.82 (0.57, 5.06)	P = 0.015
6 weeks – 6 months	Triathlon	4.93 (3.59, 6.26)	p = <0.001
	Kinemax	5.50 (3.74, 7.26)	p = <0.001
6 months – 12 months	Triathlon	1.10 (0.31, 1.89)	p = 0.007
	Kinemax	0.04 (-0.06, 1.74)	p = 0.068

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result.

Despite the significant changes notes above, independent samples t-tests of between group differences at the four assessment periods demonstrate that no differences exist between the groups at any time point (Table 3.11, Figure 3.10 and also Appendix C for statistical output).

Table 3.11 – Timed function score difference between groups at assessment

	Mean difference (95% CI)	Significance
Pre-op	1.78 (2.30, 5.86)	p = 0.39
6 weeks	-0.10 (-3.50, 3.31)	p = 0.96
6 months	0.92 (-2.19, -2.42)	p = 0.92
12 months	0.81 (-2.16, 1.70)	p = 0.81

Significance accepted at p= 0.0125 when Bonferroni correction applied. No results reach required level of significance.

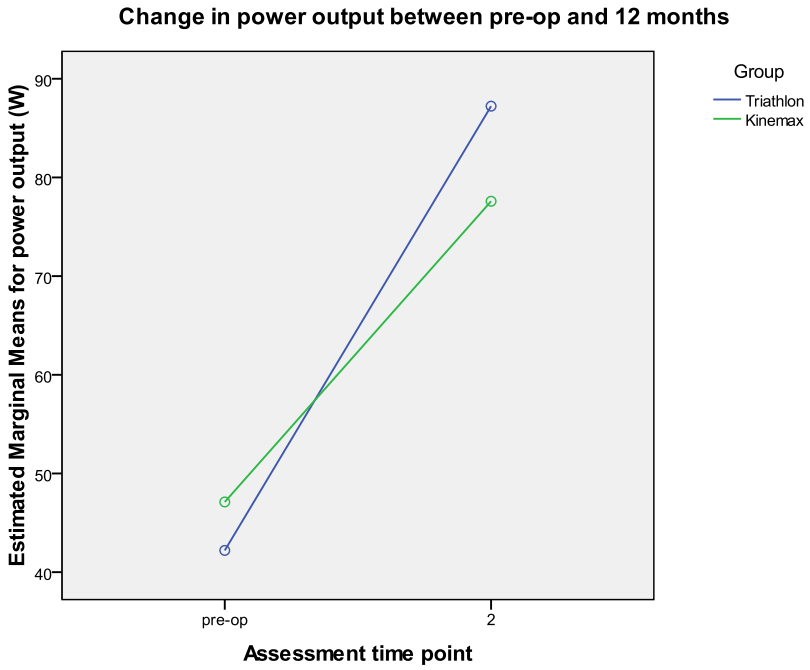
Lower Limb Power Output

RCT results: 1 year outcome

The data was normally distributed at all time points. The mean change in maximal lower limb power output (difference between pre-op and 12 month assessment periods) was 45.02 (+/- 29.33) Watts for the Triathlon group, and 30.47 (+/- 35.86) Watts for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKR surgery on the lower limb power output, $F(1, 177) = 239.98$, $p = <0.001$. There was also a significant interaction of the implant group and the recorded power output, $F(1, 177) = 8.91$, $p = 0.003$ (Figure 3.11 and also Appendix C for statistical output).

Figure 3.11 - Graphical illustration of the repeated measures ANOVA for lower limb power output. The overall increase in power following surgery is highlighted, and also the differing trajectory of change between the groups. See Figure 3.12 for errors at individual time points.



Longitudinal change across the 4 time points

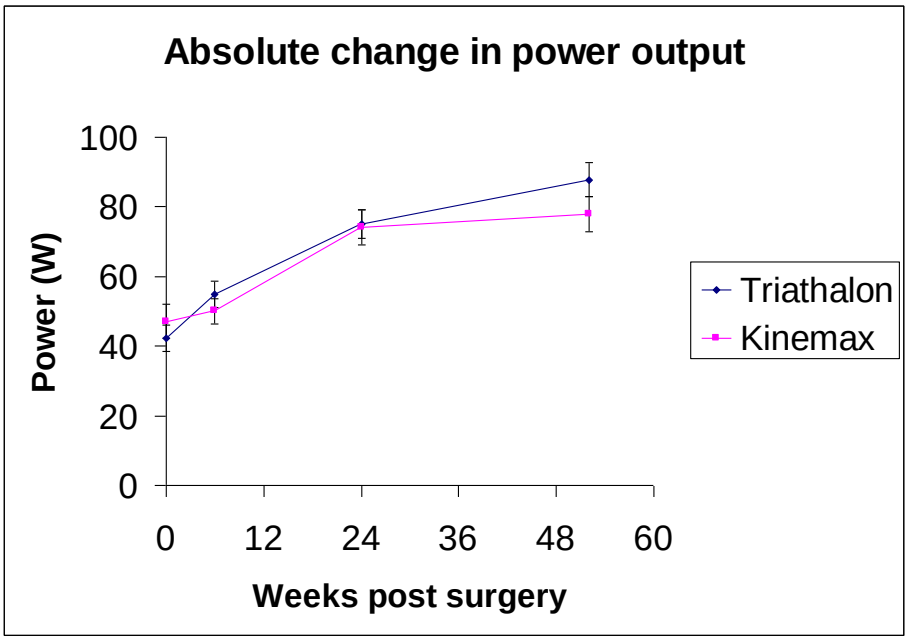
Specific analysis of the change at the four different assessment points highlights the similar overall pattern of change in power output over the 1 year testing period, but also the early (6 week) and late (12 month) differences between the two implant groups (Table 3.12 and Figure 3.12).

Table 3.12 – Power Output at assessment points

	Triathlon	Kinemax
Pre-op	42.20 (37.81)	47.11 (44.30)
6 weeks	54.86 (38.48)	50.11 (32.53)
6 months	75.14 (40.57)	74.17 (44.44)
12 months	87.84 (48.11)	77.78 (46.01)

Power output in Watts, data presented as mean (+/- SD)

Figure 3.12 - Graph demonstrates the similar overall change in power output (W) over time, but also the divergence between groups at 6 weeks and 12 months assessments.



Error bars represent the SEM.

Paired samples t-tests demonstrate significant change in power output between all time points for the Triathlon group, whereas significant improvements in the Kinemax group were observed between 6 weeks and 6 months only (Table 3.13, Figure 3.12 and also Appendix C for statistical output).

Table 3.13 – Power output change between assessments

	Implant	Mean (95% CI)	Significance
Pre-op – 6 weeks	Triathlon	-11.47 (-16.03, -6.91)	p = <0.001
	Kinemax	-1.86 (-8.68, 4.96)	p = 0.589
6 weeks – 6 months	Triathlon	-20.07 (-24.29, -15.85)	p = <0.001
	Kinemax	-25.47 (-31.54, -19.40)	p = <0.001
6 months – 12 months	Triathlon	-12.70 (-17.03, -8.37)	p = <0.001
	Kinemax	-4.51 (-10.03, 1.01)	p = 0.108

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result.

Independent samples t-tests of between group differences at the four assessment periods highlight that while the trend was for greater improvement in the Triathlon cohort, no statistically significant differences existed between the absolute power output of the groups at any individual time point (Figure 3.12 and Table 3.14 and also Appendix C for statistical output).

Table 3.14 – Power output difference between groups at assessments

	Mean difference (95% CI)	Significance
Pre-op	-4.91 (-17.24, 7.42)	p = 0.43
6 weeks	4.75 (-5.76, 15.25)	p = 0.37
6 months	0.97 (-11.77, 13.71)	p = 0.88
12 months	10.06 (-3.72, 23.84)	p = 0.15

Significance accepted at p= 0.0125 when Bonferroni correction applied. No results reach required level of significance.

Proportional lower limb power output

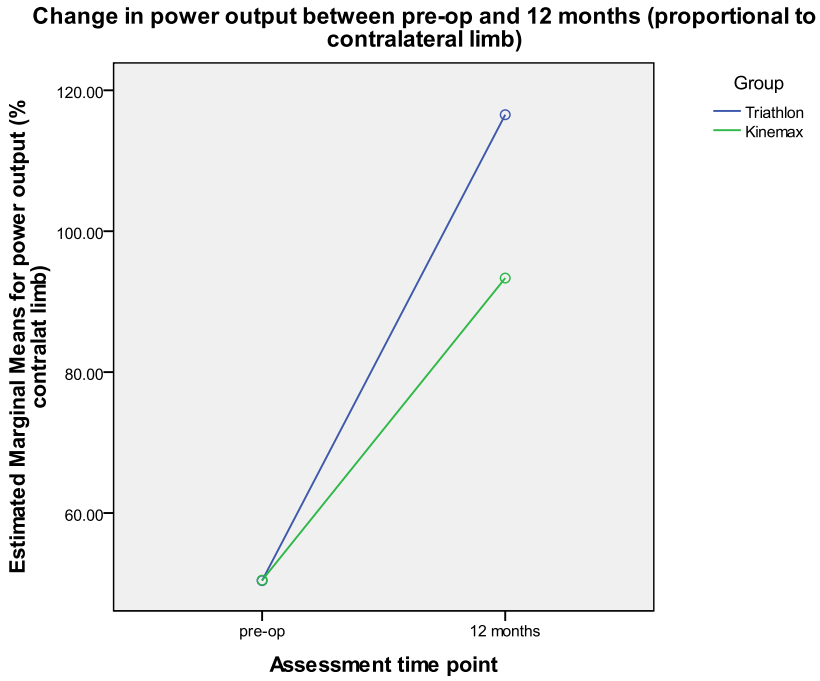
RCT results: 1 year outcome

The maximal wattage achieved by the patients at each time point was compared to the output of the contralateral limb (at 12 months) as an internal control. This data was normally distributed at all time points. The mean change in proportional lower limb power output (difference between pre-op and 12 month assessment periods) was 66.13 (+/- 44.22) for the Triathlon group, and 42.89 (+/- 47.76) for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKA on the proportional lower limb power output, $F(1, 177) = 249.09$, $p = <0.001$. There was also a significant interaction of

the implant group and the proportional power output, F (1, 177) = 11.33, p = 0.001 (Figure 3.13 and also Appendix C for statistical output).

Figure 3.13 - Graphical illustration of the repeated measures ANOVA for proportional lower limb power output. The overall increase in power as a proportion of the contralateral limb following TKA is highlighted. The differing trajectory of the two groups is also apparent, the Triathlon group achieving greater relative improvement than the Kinemax. See Figure 3.14 for errors at individual time points.



Longitudinal change across the 4 time points

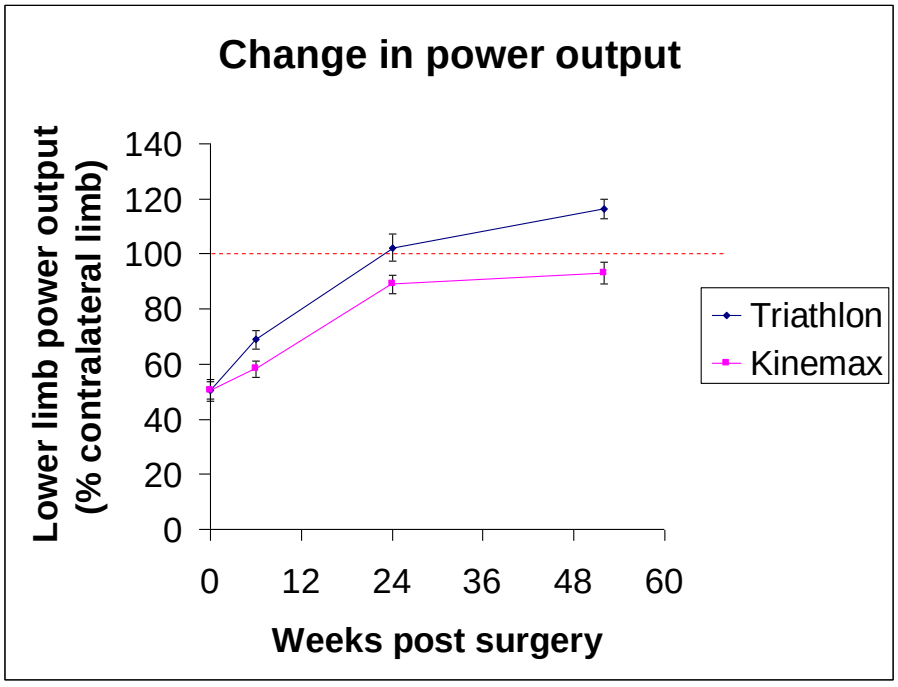
Specific analysis of the change at the four different assessment points highlights the similar overall pattern of change in power output (as a percentage of the contralateral limb) over the 1 year testing period, and also the superior improvement of the Triathlon group (Table 3.15 and Figure 3.14).

Table 3.15 – Proportional power output at assessment points

	Triathlon	Kinemax
Pre-op	50.41(32.04)	50.47 (36.67)
6 weeks	68.83 (33.30)	58.20 (27.74)
6 months	102.3 (50.13)	88.93 (30.04)
12 months	116.38 (35.57)	92.95 (35.44)

Data presented as mean (+/- SD)

Figure 3.14 - Graph of change in power output, demonstrates the similar overall pattern of change in power output over time, but also the greater proportional increase for the Triathlon group.



Error bars represent the SEM.

Paired samples t-tests demonstrate significant change in proportional power output between all time points for the Triathlon group, whereas significant improvements in the Kinemax group were observed between 6 weeks and 6 months only (Table 3.16, Figure 3.14 and also Appendix C for statistical output).

Table 4.16 – Proportional power output change between assessments

	Implant	Mean (95% CI)	Significance
Pre-op – 6 weeks	Triathlon	17.07 (24.56, -9.57)	p = <0.001
	Kinemax	6.93 (15.41, -1.55)	p = 0.108
6 weeks – 6 months	Triathlon	34.76 (43.77, 25.76)	p = <0.001
	Kinemax	30.87 (37.36, 24.38)	p = <0.001
6 months – 12 months	Triathlon	14.08 (22.42, -5.73)	p = 0.001
	Kinemax	6.03 (12.36, -0.29)	p = 0.061

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result

Independent samples t-tests of between group differences at the four assessment periods highlight that while the trend was for greater improvement in the Triathlon cohort at all assessments, only at the 12 month time point were the results statistically different between groups. The results at 6 weeks and 6 months were independently significant at p = 0.05, however did not reach the required significance when the Bonferroni correction for multiple testing was applied (Figure 3.14 and Table 3.17 and also Appendix C for statistical output).

Table 3.17 – Proportional power output difference between groups at assessments

	Mean difference (95% CI)	Significance
Pre-op	-0.06 (24.56, -9.57)	p = 0.99
6 weeks	10.63 (1.42, 19.85)	p = 0.024
6 months	13.37 (1.42, 25.33)	p = 0.029
12 months	23.42 (12.98, 33.86)	P = <0.001

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result

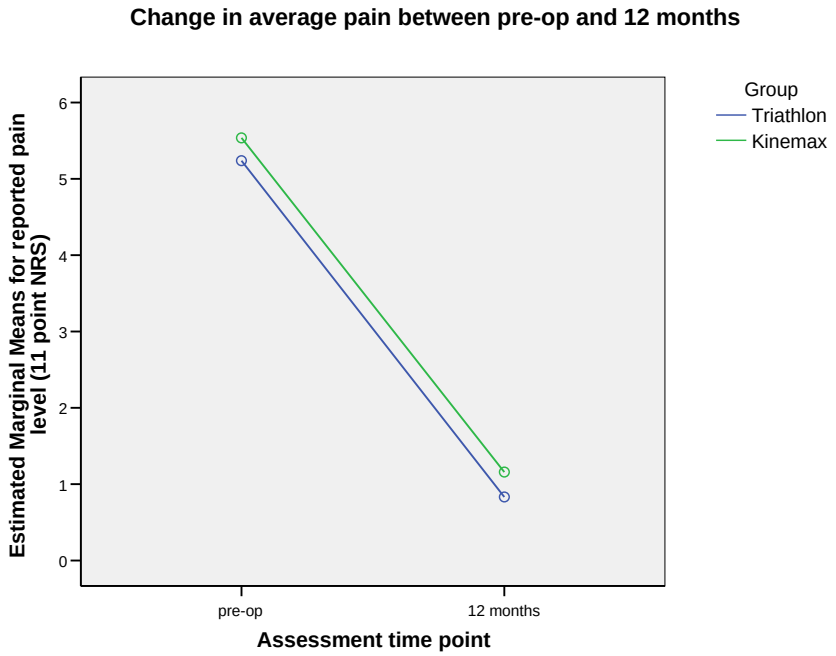
Reported average pain score

RCT results: 1 year outcome

The data was skewed to the left at 6 and 12 months post surgery. The repeated measures ANOVA considered the change in score between the assessments, and this 'change data' was normally distributed. The mean change in average pain score (difference between pre-op and 12 month assessment) was 4.41 (+/- 1.89) points for the Triathlon group, and 4.38 (+/- 2.18) for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKA on the report of average pain, $F(1, 181) = 849.9$, $p = <0.001$. There was no significant interaction of the implant group and the report of average pain, $F(1, 181) = 0.009$, $p = 0.926$ (Figure 3.15 and also Appendix C for statistical output).

Figure 3.15 - Graphical illustration of the repeated measures ANOVA for average pain score. The overall reduction on pain score following surgery is highlighted, and also the similar trajectory of change of the groups. See Figure 3.16 for individual ranges.



Longitudinal change across the 4 time points

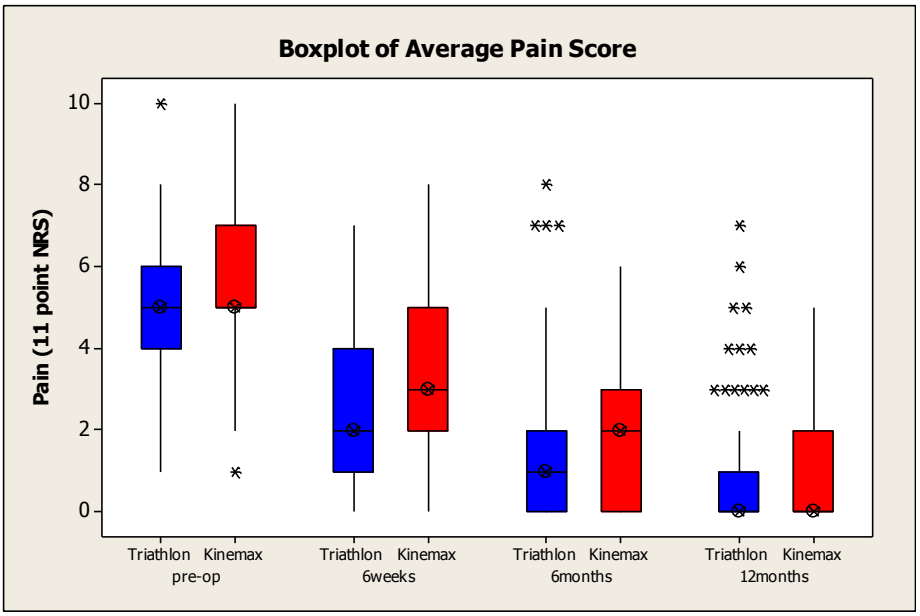
Specific analysis of the change at the four different assessment points highlights the similar overall pattern of change of reported average pain between the groups across the study timeframe (Table 3 18 and Figure 3.16).

Table 3.18 – Average pain score at assessment point

	Triathlon	Kinemax
Pre-op	5.0 (4.0, 6.0)	5.0 (5.0, 7.0)
6 weeks	2.0 (1.0, 4.0)	3.0 (2.0, 3.0)
6 months	1.0 (0.0, 2.0)	2.0 (0.0, 3.0)
12 months	0.0 (0.0, 1.0)	0.0 (0.0, 2.0)

Median pain score (+/- IQR) based on 10 point numerical rating scale

Figure 3.16 - Boxplot highlighting the similar pattern of reduction of average pain score across the assessment period. The Triathlon group appear to report less pain at all assessment points.



X denotes the median result and * an outlying data point.

Wilcoxon Signed Rank tests demonstrated statistically significant differences in average pain score between all assessment periods for both implant groups, p = <0.003 (Table 3.19 and also Appendix C for statistical output), highlighting a step wise improvement in score over the first year post-operatively.

Table 3.19 - Average pain score change between assessments

	Implant	Z statistic	Significance
Pre-op – 6 weeks	Triathlon	-7.51	P = <0.000
	Kinemax	-6.12	P = <0.000
6 weeks – 6 months	Triathlon	-6.11	P = <0.000
	Kinemax	-5.53	P = <0.000
6 months – 12 months	Triathlon	-3.80	P = <0.000
	Kinemax	-2.95	p = 0.003

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result

Mann-Whitney U-tests of between group differences at the four assessment periods confirmed that while the Triathlon scores appeared generally lower, at no specific assessment point were the reports of average pain significantly different between the implant groups (Table 3.20 and Appendix C for statistical output).

Table 3.20 – Average pain score difference between groups at assessments

	Z statistic	Significance
Pre-op	-1.30	p = 0.19
6 weeks	-1.54	P = 0.125
6 months	-1.37	p = 0.17
12 months	-1.74	p = 0.82

Significance accepted at p= 0.0125 when Bonferroni correction applied. No results reach required level of significance.

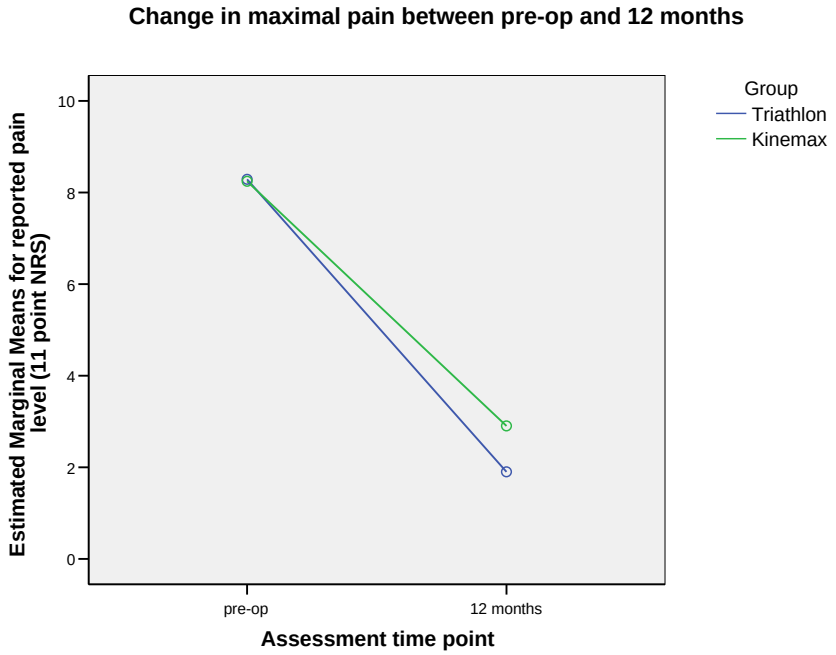
Reported maximal pain score

RCT results: 1 year outcome

The data was skewed to the left at 6 and 12 months post surgery. The repeated measures ANOVA considered the change in score between the assessments, and this 'change data' was normally distributed. The mean change in maximal pain score (difference between pre-op and 12 month assessment) was 6.39 (+/- 2.46) points for the Triathlon group, and 5.34 (+/- 2.71) points for the Kinemax group.

A two-way repeated measures ANOVA was performed to assess the overall change between pre-op and 12 month assessment, and any influence of implant group. There was a significant main effect of the TKA on maximal pain score, F (1, 181) = 1981.5, p = <0.001. There was also significant interaction of the implant group and the reported maximal pain, F (1, 181) = 7.44, p = 0.001 (Figure 3.17 and also Appendix C for statistical output).

Figure 3.17 - Graphical illustration of the repeated measures ANOVA for maximal pain. The overall reduction in pain score is apparent, and also the similar but significantly different trajectory of change between the groups. See 3.18 for ranges at individual time points



Longitudinal change across the 4 time points

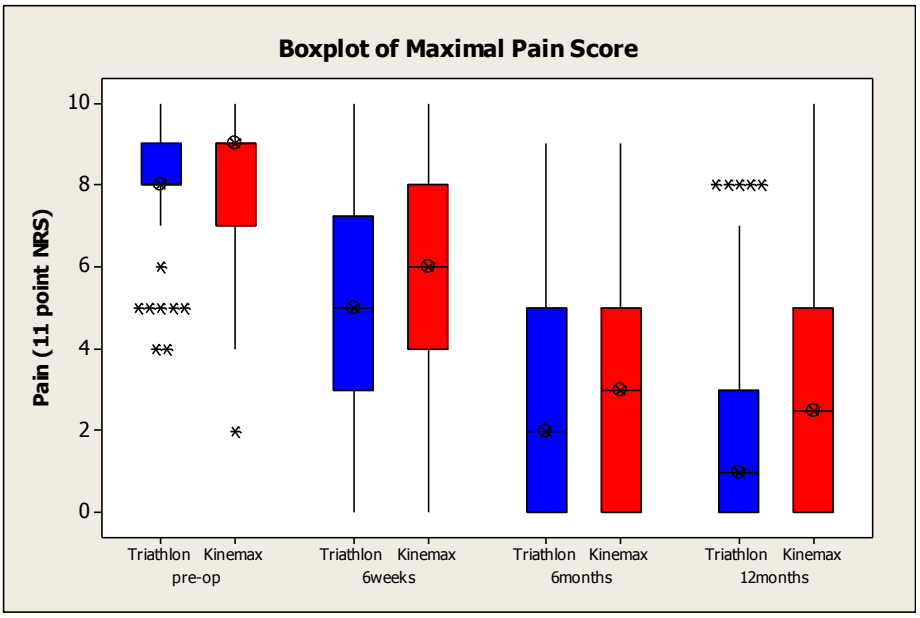
Specific analysis of the change at the four different assessment points highlights the similar overall pattern of change of reported maximal pain between the groups across the study timeframe (Table 3.21 and Figure 3.18).

Table 3.21 – Maximal pain score at assessment point

	Triathlon	Kinemax
Pre-op	8.00 (8.00, 9.00)	9.00 (7.00, 9.00)
6 weeks	5.00 (3.00, 7.25)	6.00 (4.00, 8.00)
6 months	2.00 (0.00, 5.00)	3.00 (0.00, 5.00)
12 months	1.00 (0.00, 3.00)	2.50 (0.00, 5.00)

Median pain score (+/- IQR) based on 10 point numerical rating scale

Figure 3.18 - Boxplot highlighting the broadly similar pattern of reduction of maximal pain score across the assessment period, and the reduced score of the Triathlon group at 12 months compared to the Kinemax group



X denotes the median result and * an outlying data point.

Wilcoxon Signed Rank tests demonstrated statistically significant differences in average pain score between all assessment periods for both implant groups (p = <0.001), except between 6 and 12 months in the Kinemax group (Table 3.22, Figure 3.18 and also Appendix C for statistical output).

Table 3.22 – Maximal pain score change between assessments

	Implant	Z statistic	Significance
Pre-op – 6 weeks	Triathlon	-7.56	<0.001
	Kinemax	-6.72	<0.001
6 weeks – 6 months	Triathlon	-6.76	<0.001
	Kinemax	-6.22	<0.001
6 months – 12 months	Triathlon	-3.88	<0.001
	Kinemax	-1.92	0.06

Significance accepted at p= 0.0125 when Bonferroni correction applied. Red highlighting indicates a significant result

Mann-Whitney U-tests of between group differences at the four assessment periods highlighted a similar change over the study time frame. While the overall change between groups differed (ANOVA), at no specific assessment point was there a significant difference between implant groups. Independently assessed, the Triathlon group demonstrated a lower maximal pain score at 12 months compared to the Kinemax group (Table 3.23, Figure 3.18 and Appendix C for statistical output), however due to multiple testing, this difference cannot be said to be significant.

Table 3.23 – Maximal pain score difference between groups at assessments

	Z statistic	Significance
Pre-op	0.00	1.00
6 weeks	-1.28	0.20
6 months	-0.58	0.57
12 months	-2.89	0.04

Significance accepted at p= 0.0125 when Bonferroni correction applied. No results reach required level of significance.

3.4 Discussion

It was hypothesised that single radius of curvature femoral component designs would improve patient outcomes following TKA through enhanced kinematics and superior extensor mechanism / quadriceps function. In this study, two questions were addressed; whether the implant design affected (1) overall patient outcome, and (2) specific muscle power output of the extensor mechanism. These were assessed by a comprehensive physical assessment protocol incorporating patient reported outcome and functional assessments in addition to a specific power output analysis.

Substantial treatment effects as a result of undergoing TKA were observed across the range of outcome assessments, the patients reporting much reduced pain and demonstrating enhanced functional ability 1 year post-operation. Differences were also observed between the implant groups across this time frame in OKS, maximal

pain report, flexion, extension and lower limb power output, the single radius Triathlon implant group achieving superior results in all cases.

The initial question of overall patient outcome was assessed by the comprehensive outcome assessment regime employed. The primary outcome of the RCT was change in Oxford Knee Score (between pre-operative and 12 month values), This was found to be significantly different between the groups ($p = 0.04$, Figure 3.3), the difference also reaching the accepted level of clinical significance of around 3 points (Murray et al, 2007). Of interest was that no specific differences between groups were observed in the time course of change over the year, nor were the groups significantly different at any specific time point, suggesting relative parity between groups across the post-operative time frame, in all but overall change in score.

Patient reports of pain are reported not to correlate well to nociception, the activation of sensory transduction in receptors that convey information about potential tissue damage. Von Koff et al (2000) describe pain severity as a global construct measured by both pain intensity and interference with activities. Pain intensity is a quantitative estimate of the perception of pain, or more simply, how much a person hurts. Pain affect is more complex than intensity and involves emotional arousal and changes in 'action readiness' caused by the sensory experience of pain. It is this arousal that leads to interference in daily activities.

A simple measure of pain intensity that could be measured clinically was considered the best way of assessing the change in the patient's pain between assessment periods, and the numerical rating scale the most suitable of these tools. Assessment was made of both 'worst' and 'average' pain as is recommended (Perrot et al, 2010). The assessments of function (OKS and ALF) taken as part of the trial reflect the pain affect element through its noted link to interference with daily activity. The OKS was particularly likely to provide analysis of this aspect as around half of its questions relate to pain limited function.

A suggested criticism of numerical rating scale scoring is that these scales *may* not have ratio qualities (Price et al, 1994). While this does not affect the validity, reliability or sensitivity of the scale, when evaluating the change in reported perceived pain, percentage reductions cannot be calculated as the scoring intervals may not be consistent. It has been recognised recently however that parametric statistical techniques (specifically the analysis of variance used in this analysis) are appropriate and valid when used with data that do not represent equal interval values if the number of categories on the scale in question is five or more (Philip, 1990; Van Koff et al, 2000).

The substantial reductions in pain reported in each group were broadly similar. There were no between group differences observed in average pain reported between pre-operative and 12 month post-operative assessments (ANOVA, Figure 3.15), though a significant difference emerged in report of maximal pain (ANOVA, Figure 3.17). Much of this difference can be attributed to the continued reduction in pain score in the Triathlon group between 6 and 12 months, where the corresponding Kinemax group score plateaus (Figure 3.18).

The between group difference in maximal pain reported (between pre-operative and 12 month assessment) is 1 point on the 0-10 scale. Whether this represented a clinically meaningful difference is difficult to interpret as it is debated whether or not the scale can be interpreted as interval or ratio level data. This perhaps then represents a ten percent difference between the groups in terms of change in maximal pain reported over the year, and corresponds to the superior OKS of the Triathlon group. Despite this, it must be noted that the median change in pain demonstrated over the four assessment periods (Figure 3.18) suggests no between group effect, nor is there a between group difference in either assessment of average pain report (Figures 3.15 and 3.16), suggesting little clinical relevance to this statistical finding of difference in maximal pain.

No differences are apparent between groups in timed assessment of function. The groups follow the same pattern of recovery of function post-operatively (Figure 3.10) demonstrating essentially identical change in time measured on the ALF score. Both groups demonstrate achievement of a high level of post-operative function. It may be that the ALF test is not sensitive enough to detect changes in well functioning groups and that a basement effect is limiting analysis. There is a limit to how fast these three activities can be performed when mobilising in a 'normal manner', as the patient is instructed at the onset of the test. The average times achieved of around 25 seconds for the composite assessment at 1 year are comparable to control data obtained for healthy volunteers (internal data, not displayed). Of interest is the change from pre-operative time to initial 6 week post-op assessment, where the Triathlon group demonstrates a larger improvement. This early change is independently significant, though not when assessed as part of the overall change (Table 3.11). It may be that at this specific time point in the recovery process that the Triathlon group achieved enhanced functional timed outcome, but that this comparative improvement was lost by subsequent assessments.

Range of motion was significantly improved in both groups by the TKA, and significant between group differences were observed in both flexion and extension (Figures 3.5 and 3.7). Recovery of flexion followed the same pattern for each group, but at all post-operative assessments, the Triathlon group demonstrated significantly greater flexion than the Kinemax (Table 3.5, Figure 3.6). Mean extension followed the same pattern of recovery until 6 month assessment, where the values found in the Kinemax group plateau and the Triathlon group continued to improve (Figure 3.8, Table 3.7). It is unlikely that the difference found in extension represents a clinically significant finding as it represents an average value of less than 1 degree of motion, which is below the sensitivity of the measurement tool used.

The second question, relating specifically to the power output of the extensor mechanism is perhaps the most interesting aspect of this study, as the between group differences found in lower limb power output were specifically predicted by the

implant design that incorporated a lengthened moment arm of the extensor mechanism. Substantial improvement was observed in both groups reflecting post-operative muscle recovery. There was also substantial difference between the groups, the Triathlon group demonstrating significantly enhanced change over the test period both assessed as maximal power output, and as a proportion of the contralateral limb (Figures 3.11 and 3.13).

The results expressed as a percentage of the opposite limb is the most meaningful analysis, as this acts as an internal control to quantify the power recovery. Here the same pattern of recovery was demonstrated by each group across the 4 assessment points; however the Triathlon group showed a trend of enhanced results at every assessment (Figure 3.14). Statistical significance was reached at $p<0.03$ at both 6 weeks and 6 months assessment, which does not reach the required value due to the correction for multiple testing (Table 3.17). The large differences in individual patient power output results in large confidence intervals for the mean value, which play a dominant role in the lack of between group significance detected in this study. It is likely that increased volume of data would address this discrepancy. Differences at 12 months were highly significant ($p<0.001$), the Triathlon group actually demonstrating superior power output compared to the contralateral limb, in contrast to many authors (Valtonen et al, 2009, Gapeyeva et al, 2007; Mizner et al, 2005c; Berman et al, 1991) who report typical power output being 20 percent below the opposite limb up to 12 months post-op.

Substantial differences in lower limb power and flexion were observed between the implant groups. A three point change in the OKS and small difference in maximal pain reported were also apparent. Interestingly, this did not seem to substantially affect the ability of the patients to perform timed functional tasks. That range of motion did not affect function supports the results of Nutton et al (2008) who demonstrated that high flexion implants do not to make any difference to functional ranges of flexion. Power output though is known to relate strongly to functional ability, particularly to tests of walking and climbing the stairs (Lamb and Frost,

2003, Mizner et al, 2005c). It is therefore interesting that the differences apparent in power output did not translate to the ALF timed assessments. This supports the theory that a basement effect was reached in this assessment, and that a well performing total population limits the analysis of this particular assessment.

Strengths and limitations

This is the only assessment of post-operative function directly comparing single and multi-radius implants that has both the statistical power and range of assessment tools to assess patient functional outcome comprehensively. Further, this is the only report to assess the functional outcome of the Triathlon prosthesis in such a manner. A particular strength is the double blind and randomised conditions of the trial.

Covariates were not specifically addressed as the randomised nature of the trial limits any such effect. The trial is of a pragmatic design to best reflect the typical presentation of knee arthroplasty patients pre-operatively. While the inclusion of patients with other joints that suffer from osteoarthritis or have had previous arthroplasty may theoretically affect subsequent function, the exclusion criteria was set to prevent recruiting patients whom it was felt suffered functional limitation due to these other joints, and again the randomised nature of the trial limited any specific effect on the between group differences. The overall function of both groups can be reported as excellent, and it is unlikely that any such influences affected the mean post-operative function of either group.

All surgical and post-operative factors were kept identical as far as possible. Routine surgical procedures and post-operative care were not affected by recruitment to the trial, and the researcher performed all post-operative clinical assessments adhering to local clinical protocol blinded to the underlying prosthesis. Other implant factors that are thought to impact outcome such as bearing surface and cruciate retaining design were also identical.

The lack of clarity as to which outcome measures should be employed to investigate research of this nature in clinical populations demanded a pragmatic and comprehensive trial design, thus a blinded randomised controlled trial was chosen. A particular problem when designing this study was the lack of information as to the relative importance, and indeed relationship between, patient report measures of outcome and performance measures of outcome. The data collected in this study will be analysed to allow comment on these issues (and is presented in Chapter 4). Had this information been available prior to the development of this trial, it may have driven the specific design and outcome measures employed. The absence promoted the most pragmatic trial design that would confer information through various assessments as the most appropriate methodology to assess the question of power output, and also put that information into context within the wider sphere of patient outcome. This has been achieved, and information is now available to further assess the relationship between different types of outcome assessment tools.

Alternative trial designs, such as a cohort study, utilising additional specific testing of the quadriceps muscle group by means of perhaps isokinetic dynamometry and imaging modalities (to assess the muscle cross-sectional area) could also be employed to investigate the specific differences pertaining to the extensor mechanism muscle function. It was not possible to include such intensive analysis tools within the context of the study that was performed, due to the multi-modal assessment protocol that was required. Additionally clinical time constraints and the size of the population assessed limited such detailed evaluation. An additional cohort study may provide advantageous supplementary information by which to assess the specific issue of the muscle function facilitated by these differing implant designs.

Conducting a cohort study of a specific population group with a restrictive inclusion criteria may hypothetically have reduced the standard deviation of the power output scores, aiding interpretation of the between group results. In terms of the hierarchy of evidence however, the RCT is a superior methodology to the cohort study as it minimises the chance of bias through the randomisation process. This limits the

chance of (previously unknown) confounding variables differing between the groups. From a statistical perspective, the randomised nature of the trial conducted justifies the pragmatic inclusion criteria. The baseline power data was found to be comparable in each arm of the trial, and was strikingly similar when assessed against the internal control of the contralateral limb. This strengthens the conclusions of this chapter regarding the influence of the implant design on the subsequent power output.

The Kinemax® total knee replacement is an older implant of 2nd generation design, while the Triathlon® total knee replacement is of 4th generation lineage. The major difference that distinguishes these implants is that they are based on alternative ideas regarding the underlying kinematics of the knee. Further differences do however exist between these implants that could in principle also affect subsequent patient outcome. Additional design changes incorporated into the Triathlon® implant that differ form the Kinemax® include a 'sided' patello-femoral grove, decreased size but increased angle of anterior flange (reducing the patello-femoral bulk of the prosthesis) and a shortened posterior condylar offset (These changes are visualised for clarity in Appendix I). The changes to the anterior aspect of the implant may influence such factors as anterior knee pain, and the posterior condylar offset the flexion achieved (Nutton et al, 2008). A third arm comparing another, more modern, multi-radius implant that incorporates some of these additional design changes would have been beneficial to address the extent to which these additional factors influenced the overall patient outcome.

It was necessary to compare the new Triathlon® implant with an established current implant that demonstrated 10 year survival to benchmark the new design in accordance with best practice (BOA / BASK, 2010). The Kinemax® knee replacement was a good example of a successful 'condylar' multi-radius implant (Wright et al, 2004; Back et al, 2001), further Pradhan et al (2006) note similar survival curves and generally good 10 year performance of all the condylar devises (Kinemax, PFC, AMK and Total Condylar) used at their unit, in a report of over 3300 cases.

In conclusion, this double blind randomised controlled trial demonstrated functional differences in overall patient outcome based on the implant used at time of surgery. The modern implant design allowed enhanced patient function as observed by the change in OKS and an additional raft of outcome assessment tools.

The most striking difference detected between the implant groups was in the lower limb power output of the patients, a factor specifically predicted by the axis of rotation related design changes through an enhanced moment arm of the extensor mechanism. Support can be lent to the single radius design improving quadriceps function post-operatively through mechanically reducing the work the quadriceps muscle need perform to extend the knee. The wider effect of this difference in quadriceps power and further relationship between patient reported outcome and direct assessment of outcome is developed in the following chapter.

4 The Role of Patient Report and Performance Assessment in Determining Outcome Following TKA

4.1 Introduction

Assessment of outcome has shifted away from the dichotomous criteria of success or failure and towards the use of quantifiable outcome measures. The use of patient reported outcome measures (PROMs) to assess functional outcome and quality of life following TKA has now been adopted as a requirement of all providers of elective arthroplasty surgery of NHS patients (DoH, 2008). Advocates of these patient reported tools suggest that they provide a sophisticated measure of whether a treatment has worked in terms of how much better a patient feels (Timmins, 2008). Terwee et al (2006) define a performance measure as one in which the individual is asked to perform an activity that is evaluated in a standardised manner using predefined criteria such as time taken. Self report assessments are ones in which individuals are asked to report their perceived level of functioning during daily activities described in standardised questions.

Recently a small number of authors have suggested that PROMs alone are not sufficient to assess patient outcome. Mizner et al (2011, published ahead of print) argue that patient report measures fail to capture the actual changes in functional performance after TKA. Performance assessments are additionally required to obtain the full picture of the patient's physical function (Stratford and Kennedy, 2006). The benefits and limitations of each type of assessment have been discussed in Chapter 2, though it is specifically suggested that the patient report measures have low construct validity, and that this is perhaps due to an influence of pain (Terwee et al, 2006; Stratford et al 2006). Despite this, Bream et al (2009) claim that patients are the 'gold standard' judges of symptoms and quality of life, and that PROMs thus provide an accurate indication of the outcome of surgery.

Scant analysis has been conducted directly comparing patient report and performance data, and only one recent study (Mizner et al, 2011, published ahead of print) has assessed how this relationship changes over time following TKA. The Mizner et al paper reports poor concurrent validity between the assessment types and that patient report measures overstate the functional ability of the patient, particularly in the acute post-operative phase. These authors compared the responsiveness of the two types of outcome assessment over time, but did not investigate the relationship between assessments at individual time points.

Beard et al (2010) pragmatically note that the choice of outcome assessment tool will depend on a variety of factors relating to the aim of the assessment, its context and the level of detail required. It remains somewhat unclear as to the most appropriate method of examining function following TKA, and how to interpret the results. The association between patient report and performance data has not been previously investigated in sufficiently large volume to answer these questions.

A large volume of performance data was presented in the previous chapter, prospectively collected in tandem with the patient reported Oxford Knee Score. The separate performance measures and pain assessments are examined against the patients' report of pain and function via the OKS with the aim of assessing the association between the two types of measurement at the different time points.

It is hypothesised that correlation will be apparent between the OKS and the additional pain and performance measures and further that the variation in OKS would be explained by variation in the additional pain and functional measurements.

4.2 Methods

The patient outcome data recorded as part of the RCT presented in the previous chapter was assessed to determine the relationship between patient report outcome measures and direct assessment of patient physical performance. This data was recorded for patients listed for TKA pre-operatively and at 6 weeks, 6 months and 12 months post-operatively.

The Oxford Knee Score is a widely employed patient derived outcome questionnaire (KAT trial group, 2009; Davies, 2002). It consists of 12 equally weighted questions addressing pain and functional activity, each scoring form 1 – 5. Best possible scores are 12 and worst 60. It is a joint specific instrument that was developed to assess the outcome of knees in randomised trials and has undergone rigorous assessment of reliability, validity and responsiveness. Specific attempt was made during the construction of the OKS to minimise the influence of comorbidities (Murray et al, 2007).

The additional physical performance assessments were the Aggregated Locomotor Function test, lower limb power output (Leg Extensor Power Rig) and goniometric assessment of knee flexion. For the pain component, both the 11 point (0-10) numerical rating scales for average and maximal pain experienced were assessed. The individual assessment tools have been described previously (Methods section, Chapter 3).

The additional pain and physical outcome assessments were compared against the patient report of these parameters (OKS) to determine the extent of the variance of the patient report that the direct assessments were able to explain.

Statistical analysis

Data was analysed using the Minitab (release 15) software. Differences in outcome at each assessment between the outcome measures were assessed with paired samples t-tests.

Correlation of performance variables with the OKS was assessed, and regression analysis performed on those that formed significant associations. Multiple linear regression analysis was employed using a stepwise model building technique to screen out predictors not associated with the response. Significance was accepted at $p = 0.05$. The adjusted R^2 value (adjusted for the number of predictors in the model) and Mallows Cp statistic (to assess how well the model fits the data) are reported to allow comparison between models at different time points.

4.3 Results

Data for 183 cases were available for analysis. Mean age of the patients was 68.4 (9.03) years, the gender split was 71 male (38.8%) to 112 female patients (61.2%). Table 4.1 displays the mean score for each of the outcome assessments at each time point.

Table 4.1 – Mean outcome scores for all assessment time periods

	Pre-op	6 weeks	6 months	12 months
OKS	40.87 (7.519)	32.58 (8.99)	24.97 (8.24)	22.17 (7.90)
Flexion	104.70 (14.62)	96.39 (13.22)	104.27 (12.35)	107.30 (11.76)
ALF	35.38 (14.24)	31.64 (11.51)	26.55 (7.77)	25.61 (6.60)
Power	44.40 (40.79)	52.72 (35.91)	74.72 (42.18)	83.33 (47.32)
Pain (Max)	8.27 (1.46)	5.34 (2.53)	3.00 (2.70)	2.35 (2.57)
Pain (Ave)	5.37 (1.54)	2.97 (2.01)	1.52 (1.80)	0.98 (1.53)

Mean (SD) scores are displayed for all variables. OKS = Oxford Knee Score, ALF = Aggregated Locomotor Function score (seconds), Power output (watts), Maximal and Average Pain scores (0-10 NRS).

All variables, changed significantly between each assessment time point (paired samples t-test, p = < 0.001 in all cases). This change was almost universally positive in nature, the only exception being a reduction in flexion from pre-operative to 6 week post-operative assessment.

Relationship of OKS with pain and functional measures over time

Correlation coefficients are displayed in table 4.2 for the relationship between Oxford Knee Score and the functional parameters at all assessments. Highly significant correlations (p = <0.001) were found between the OKS and all other assessments. Generally, high levels of correlation (r = 0.6 - 0.7) were seen between the OKS and pain reports post operatively, though these were less well associated pre-operatively (r = 0.4). Functional parameters (ALF and power output) were modestly associated (r = 0.4) with OKS, and measure of knee flexion the least well correlated variable (r = 0.3).

Table 4.2 – Correlation of OKS with pain and performance variables

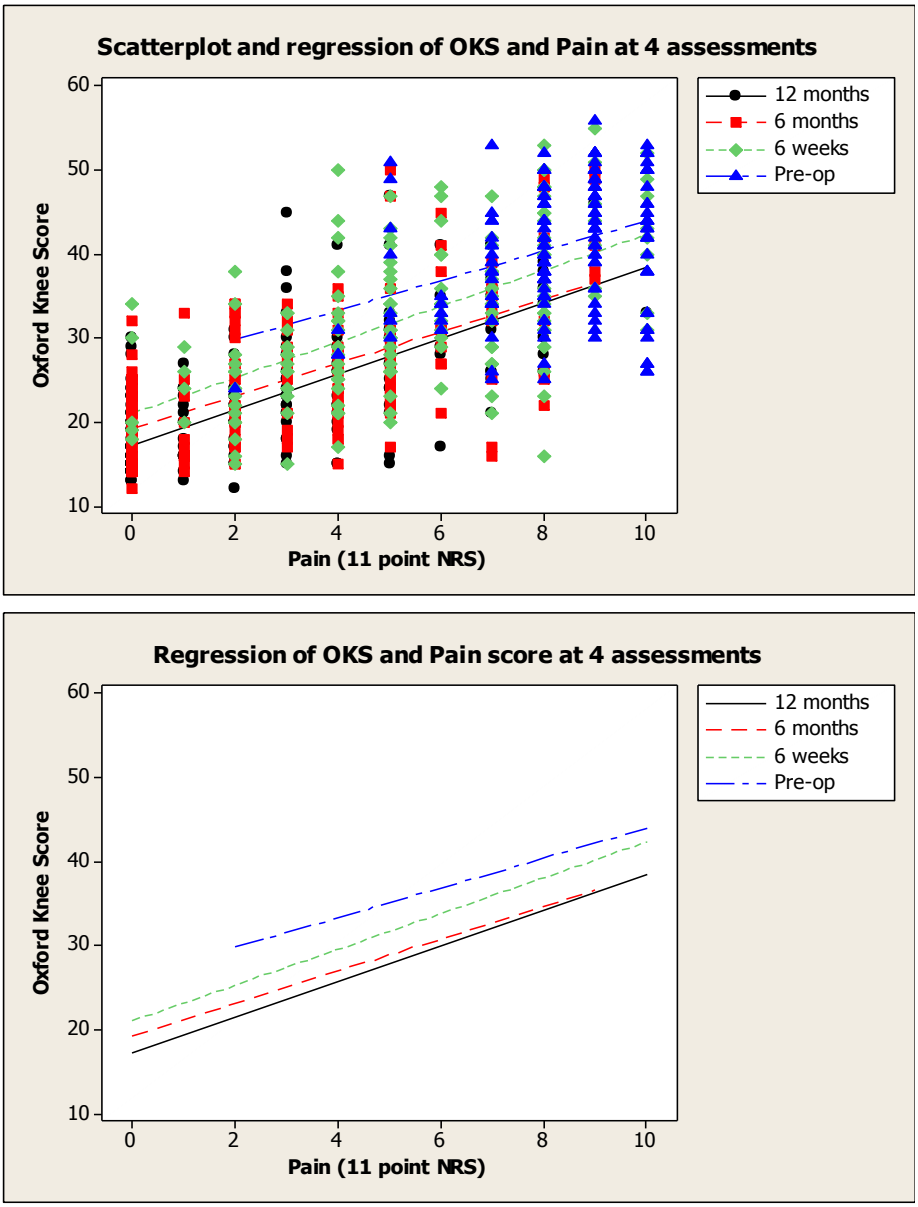
	Pre-op	6 weeks	6 months	12 months
OKS and ALF	0.31	0.41	0.54	0.49
OKS and Power	0.39	0.37	0.37	0.39
OKS and Flexion	0.36	0.27	0.25	0.25
OKS and Average Pain	0.44	0.60	0.63	0.70
OKS and Maximal Pain	0.35	0.60	0.61	0.65

Individual correlation coefficient r values are displayed for each variable at each assessment time point, p = <0.001 in all cases

Pre-operatively, both pain and function (ALF) assessments were less well correlated to OKS than they are post operatively. The relationship with power remain stable both pre and post-operatively. Post-operative correlation between the OKS and both power and pain was relatively stable across the assessments, suggesting a similar magnitude of improvement in both the OKS and these measures over this time period

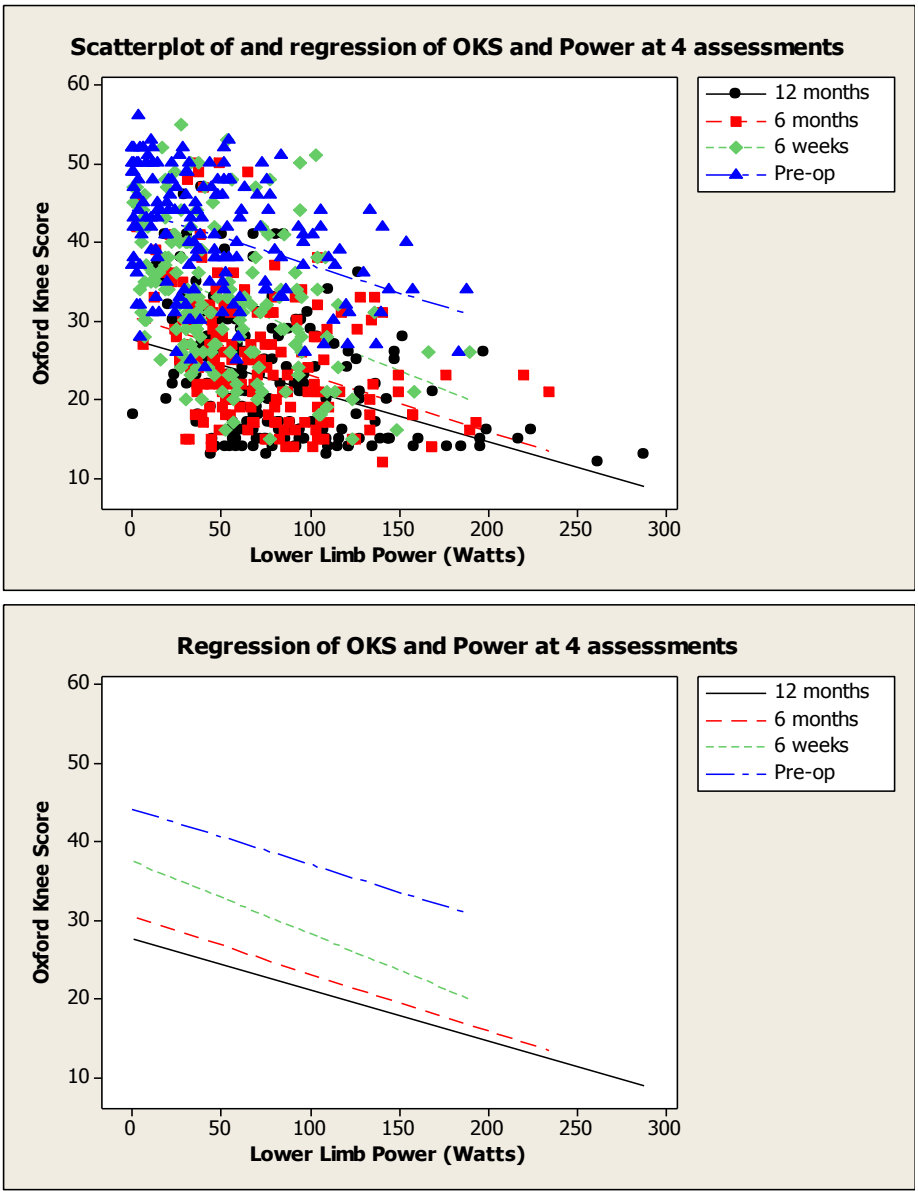
(Figure 4.1 and 4.2). Correlations with ALF score however improve over time (Figure 4.3).

Figure 4.1 - Relationship between Oxford Knee Score and pain report. Figure demonstrates the linear relationship between maximal pain report and the OKS at all 4 assessments. The top figure displays the regression on the underlying scatter plot of the data, the bottom figure the individual regression lines for clarity. The gradient of the regression lines are comparable at each post-operative assessment, though follow a different trajectory pre-operatively.



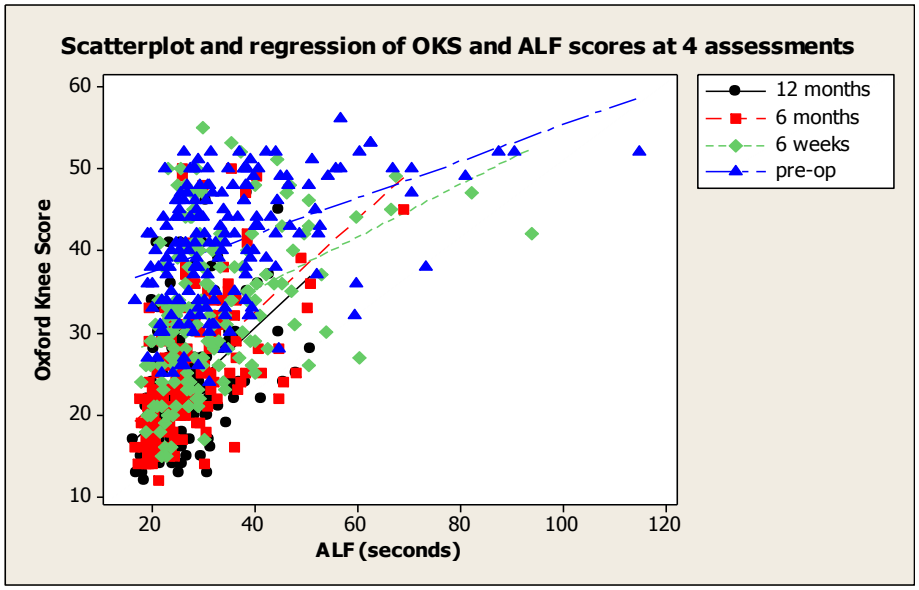
Correlation coefficient values are presented in Table 4.2.

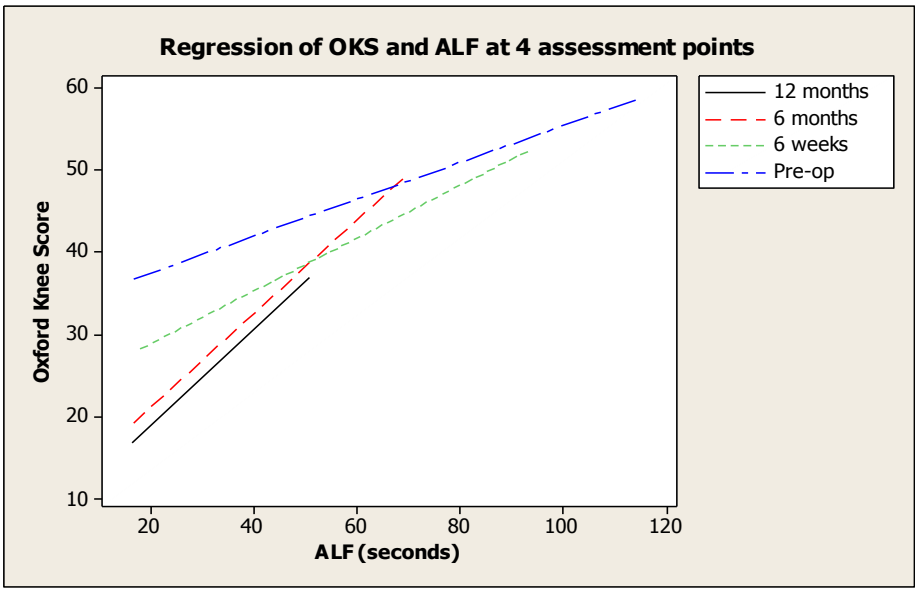
Figure 4.2 - Relationship between Oxford Knee score and power output. Figure demonstrates the linear relationship between maximal lower limb power output and the OKS at all 4 assessments. The top figure displays the regression on the underlying scatter plot of the data, the bottom figure the individual regression lines for clarity. The gradient of the regression lines are broadly comparable at each assessment.



Correlation coefficient values are presented in Table 4.2.

Figure 4.3 - Relationship between Oxford Knee Score and functional assessment. Figure demonstrates the linear relationship between functional assessment (ALF) time and the OKS at all 4 assessments. The top figure displays the regression on the underlying scatter plot of the data, the bottom figure the individual regression lines for clarity. The gradient of the regression lines is similar at 6 and 12 months post-operatively, but notably different at the other assessments, highlighting the differing relationship between OKS and ALF score at the different assessment time points.





Correlation coefficients are presented in Table 4.2.

Pain was less well correlated with the functional assessments (ALF and power output) than it is with the OKS, these correlations are displayed in table 4.3. Pain correlated consistently poorly with both ALF and Power assessments at all assessments (r = 0.15 - 0.30). The power and ALF scores were modestly correlated (r = 0.5) and this relationship remains stable throughout the assessments. The separate measures of pain correlate highly in the first three assessments (r = 0.8) though modestly at 12 month follow-up. Comparative decrease in both maximal and average pain was seen at earlier assessments (Table 4.1). The treatment effect of the TKA results that many patients report no pain at 12 months. This affected the resultant correlation between pain assessments. Despite high degree of correlation between average and maximal pain report, the maximal report of pain does not correlate with either power or ALF assessment pre-operatively, though average report of pain does so poorly.

Table 4.3 – Correlation of pain and performance variables

	Pre-op	6 weeks	6 months	12 months
ALF and Average Pain	0.17 (p=0.02)	0.27	0.27	0.25
ALF and Maximal Pain	0.03 **(p=0.63)**	0.26	0.25	0.25
Power and Average Pain	0.30	0.30	0.21	0.21
Power and Maximal Pain	0.06 **(p=0.47)**	0.30	0.15 (p=0.05)	0.19
ALF and Power	0.49	0.55	0.48	0.48
Average and Maximal Pain	0.76	0.83	0.75	0.49

Correlation coefficient r value is displayed, p = <0.001 unless stated. Non significant results highlighted in bold type.

Multiple linear regression demonstrated that the results of the pain and functional assessments explain much of the variance in OKS post-operatively. Functional measures, pain assessment and demographic variables were assessed as to their relevance to the OKS outcome; the results are displayed in table 4.4. Only pain measures, ALF and power assessments were correlated at all time points.

Stepwise regression modelling was performed to screen out predictors not associated with the response variable (OKS) based on the criteria of alpha (to enter or remove from the model) of 0.05. Results are displayed in Table 4.4.

Table 4.4 – Multiple linear regression analysis for OKS at each assessment

	Pre-op	6 weeks	6 months	12 months
Flexion	0.069ns	0.057ns	0.671ns	0.273ns
ALF	<0.000	0.005	<0.000	<0.000
Max Pain	0.017	0.005	0.001	<0.000
Ave pain	0.006	0.001	0.055	<0.004
Power	0.022	0.020	0.015	0.003
Gender	0.225ns	0.071ns	0.228ns	0.210ns
Age	0.184ns	0.027	0.004	0.254ns
Stepwise model				
Adjusted r^2	34.8	44.0	56.9	62.3
Mallows Cp	8.3	10.7	7.3	6.4

Multi-variant regression analysis for OKS at each time point, p-value reported for individual factors, significance accepted at p = 0.05, non significant results noted (ns). Stepwise regression models were constructed. Regression coefficients are displayed for these at each time point. Full data output displayed in Appendix D.

At 12 months, the regression model included maximal pain, ALF, average pain and power output (explaining 62.3% of the variation in OKS). At 6 months, the model consisted maximal pain, ALF, age and power output (explaining 56.9%). At 6 weeks the model consisted of maximal pain, ALF and average pain (explaining 44.0%). Preoperatively the model consisted average pain, ALF, maximal pain and flexion (explaining 34.8%). See Appendix D for statistical data output.

Of all the assessments conducted, pain appeared to be the dominant factor in the variation of the OKS at every assessment. Table 4.5 displays the contribution of the factors that best explained the variation in OKS; pain report (maximal and average), ALF score and power output. The regression best explained the variation in OKS at all time points, however the report of pain demonstrates values only slightly less than

those of the multiple response analysis, again highlighting its stronger association with the OKS compared to the functional assessments.

Table 4.5 – Comparison of regression model with pain and function values

	Regression model	Pain reports	ALF score	Power output
Pre-op	34.8	20.4	17.9	14.4
6 weeks	44.0	40.0	15.9	13.4
6 months	56.9	41.4	29.2	13.2
12 months	62.3	51.1	23.8	15.0

Adjusted R^2 values are displayed for both multiple and simple regressions to allow direct comparison.

Pre-operative predictors of 1 year patient outcome

Multiple linear regression analysis and then stepwise regression modelling was performed (alpha to enter or remove from the model of 0.05) to assess the pre-operative predictors of post operative OKS. Results are displayed in table 4.6, see Appendix D for statistical data output.

The pre-operative assessments are poor predictors of post-operative patient reported outcome. Table 4.6 highlights this poor relationship. Only pre-operative OKS was significantly related to post-operative values at any time point. This score predicted around 20% of the variance at 6 weeks, 17% at 6 months and 8% at 1 year.

Table 4.6 – Pre-operative predictors of post-operative OKS

Pre-op predictor	6 weeks	6 months	12 months
Flexion	0.528ns	0.469ns	0.563ns
OKS	<0.001	<0.001	0.028
ALF	0.043	0.134ns	0.533ns
Power	0.703ns	0.533ns	0.341ns
Pain (Max)	0.827ns	0.962ns	0.457ns
Pain (Ave)	0.807ns	0.134ns	0.220ns
Gender	0.725ns	0.509ns	0.162ns
Age	0.423ns	0.162ns	0.536ns
Stepwise model			
R-Sq (adjusted)	20.4	17.1	8.2
Mallows Cp	0.3	1.1	2.6

Multiple regression for 12 month OKS using pre-operative variables. p-values reported for individual factors, significance accepted at p = 0.05, non significant results noted (ns). Stepwise regression models were constructed. Regression coefficients are displayed for these at each time point. Full data output displayed in Appendix D.

4.4 Discussion

Recently the relationship between direct assessment of function and patient reported function has been questioned, with some authors claiming PROMs are inadequate to assess function, and that performance data is required to supplement them. Little prospective patient report and performance data has been concurrently collected with which to draw conclusions as to the relationship between the two types of assessment. No previous study has used multiple linear regression analysis to compare patient report and performance data at multiple time points.

This study analysed 183 datasets pre-operatively and at a further 3 assessments in the year following TKA. The highly validated and well utilised OKS that assesses patients' pain and function was compared with direct measurements of these attributes. In addition to the composite ALF timed test of function, specific power output and degree of knee flexion were also evaluated as to their effect on patient reported outcome.

As expected, significant correlations were found between the Oxford Knee Score and separate specific assessments of pain and functional outcome. Additionally highly significant correlations between OKS and both lower limb power output and knee flexion were also observed (Table 4.2).

The relationship between OKS and separate pain and functional measures over the assessment period is displayed in the Figures 4.1, and 4.3. The good relationship between patient report of pain (NRS) and patient report of joint specific pain and function (via the Oxford Knee Score) remains similar throughout the assessment period, though the correlation is weaker pre-operatively (Figure 4.1). The changing relationship between timed assessment of function and OKS is displayed in figure 4.3. This correlation between functional assessment and report of function improved over the rehabilitative period, and coincides with substantial reductions in reported pain post-operatively (Table 4.1).

It has been suggested that pain interferes with patient report of function and these results provide support for this assessment. Stratford and Kennedy (2006) examined the association between performance rated assessments of pain, functional tests and the WOMAC self report of pain and physical function in 85 patients suffering from osteoarthritis. Interestingly, they found pain to be the strongest determinant of the WOMAC function sub-score, and change in pain the strongest determinant of change in functional score. They used this data to suggest that patients report their *experience* of performing an activity, while the *ability* to perform the activity is assessed by the performance measure.

The data presented in this chapter provided support for this theory. Pain scores were particularly well associated with the OKS while timed functional ability was modestly associated. Pain is poorly correlated with composite functional activity time achieved ($r = 0.25$, Table 4.3). This was unexpected as pain is known to limit physical function, particularly in the case of osteoarthritis of the knee. While a

significant association is apparent, the proportion of variation explained is only of the order of 6%.

It is notable that as the effect of pain is reduced in the year following surgery, the association between ALF and OKS improves. The theory of pain levels affecting reported level of functional ability fits well with the improving relationship of reported function and direct assessment of that function as the pain level experienced reduces post-operatively.

It has been well documented that range of motion does not affect functional outcome (Nutton et al, 2008), and thus unsurprising that this was confirmed here. That the maximal lower limb power output did not interact strongly with either OKS or reports of pain is however interesting. Previous authors have demonstrated the link between pre-operative strength and post-operative function (Faulkner et al, 2010; Mizner et al, 2005b; Lingard et al, 2004; Lamb and Frost, 2003). The association between power and function is confirmed in this study, through modest correlation with both ALF score ($r = 0.5$) and OKS ($r = 0.4$) at all assessment points (Table 4.2 and 4.3). The striking stability of both these relationships at all 4 assessment times confirms the link between power and function, but also highlights the differing relationship found between the objective timed functional assessment (ALF) and the OKS.

Surprisingly, pain does not seem to influence the maximal power output achieved as poor correlation of these variables was found (Table 4.3). Further, maximal pain levels reported pre-operatively had no interaction with the ability to generate force. The power output is an objective test that results in ratio level data. This therefore suggests that any error in measurement is likely to be attributable to the patients' report of pain. It is very difficult to rate the level of pain one is experiencing objectively as it is linked to prior experience of pain, attitude towards pain and coping strategies developed to deal with it. It must be assumed that patients try to

report this pain as best they can and that it is the complexity of the constituents of the pain experience that clouds the report.

Pre-operative patient report of maximal pain did not correlate with either power or ALF score. This may reflect an inaccurate report of this constituent of pain intensity and behaviour. At the same time point, report of average pain does correlate with both physical measures albeit poorly. Interestingly average and maximal pain reports were found to correlate highly to each other pre-operatively. It may be that these differences found highlight the differing components of pre-operative pain reported by the two separate measures. This lends support for a more rigorous assessment of patient pain report than can be achieved by a single average of overall pain.

Change over time

The results demonstrate (with the notable exception of range of knee flexion) that all aspects assessed significantly improved from pre-operative values by the initial six week assessment, and then subsequently improved further by six months and again by one year assessments (Table 4.1). This is in direct contrast to the only other study that has assessed how the PROMs / physical performance relationship changes over time (Mizner et al, 2011 published ahead of print).

Mizner et al (2011, published ahead of print) assessed the relationship between patient report of function (via SF36 questionnaire) and various performance measures in 100 patients prior to TKA and at one and twelve months post-operation. They report that physical measures provided a more comprehensive perspective of recovery after TKA as the patient report measures did not reflect the acute worsening of scores seen in the performance measures at four week assessment. Patients in this study actually reported improvements on the SF36 questionnaire where deficits were observed in functional tasks (muscle weakness, timed-up-and-go test and walking distance).

Mizner et al. use this information to criticise the 'over-estimation' of patient report data in the acute post-operative phase. It may be that the two week difference n assessment time between this study and that of Mizner et al. is highly important in the recovery following TKA. The Mizner group are the same authors that vigorously promote the idea of failure in voluntary activation of the quadriceps muscle in the early post-operative phase accounting for much of the poor functional performance post-operatively. This, and potentially a difference in the post-surgical pain experienced, having reduced further in the extra two weeks post-operatively may help explain why the results presented here differ so markedly.

Stratford and Kenndey (2006) note that correlations of the value r = 0.4 to 0.6 between performance and self report assessments have been used to argue that a common attribute is being assessed. Correlations however can only assess whether a relationship exists between two variables, only regression analysis assesses the extent to which variation in the response variable is explained by variation in the predictor variables. Regression analysis of the larger of these suggested correlations would result in R^2 values of around 0.36, i.e. that one variable explains' around a third of the variation in the other, and not that a common attribute is being assessed.

The regression analysis performed with this data demonstrates that the pain component accounts for the largest proportion of variation in the OKS. Multiple regression models were created using the stepwise modelling approach with the addition of the ALF score, though few other variables were relevant (Table 4.4 and Appendix D).

Around 60% of the variation in OKS was explained by assessment of pain and performance variables at six and twelve months. The poor pre-op relationship of around 35% variation further highlights the inconsistent relationship between the patient reported data and that directly assessed, possibly due again to the pain influence (highest levels are reported pre-operatively, Table 4.1) as suggested.

Even the best regression equations derived fail to explain around 40% of the variation in OKS post-op. Many other factors are known to influence post-operative outcome, in this case gender did not, age did at the early assessments, specific co-morbidities or influences of other joints were not assessed, though all data is drawn from the RCT reported in the previous chapter that had inclusion criteria relating to reasonable physical function and will have diluted any such effect. Further the factors relating to the centre and surgeon volume are not likely to be relevant in this case as all operations were conducted at the Royal Infirmary of Edinburgh by experienced consultants, and technically performed in the same manner as part of the randomised trial.

The pre-operative functional and pain measures were unable to explain any of the variance in post-operative OKS at any time point. Only pre-operative OKS displayed any effect in predicting post-op score. Interestingly, the relationship was strongest at early assessment and diluted with time post-operatively (Table 4.6). This was an unexpected finding and somewhat at odds with the accepted understanding that pre-operative function should be considered as a co-variable in analysis of change in function. It highlights the widely varying levels of patient outcome and the complexity of trying to analyse this.

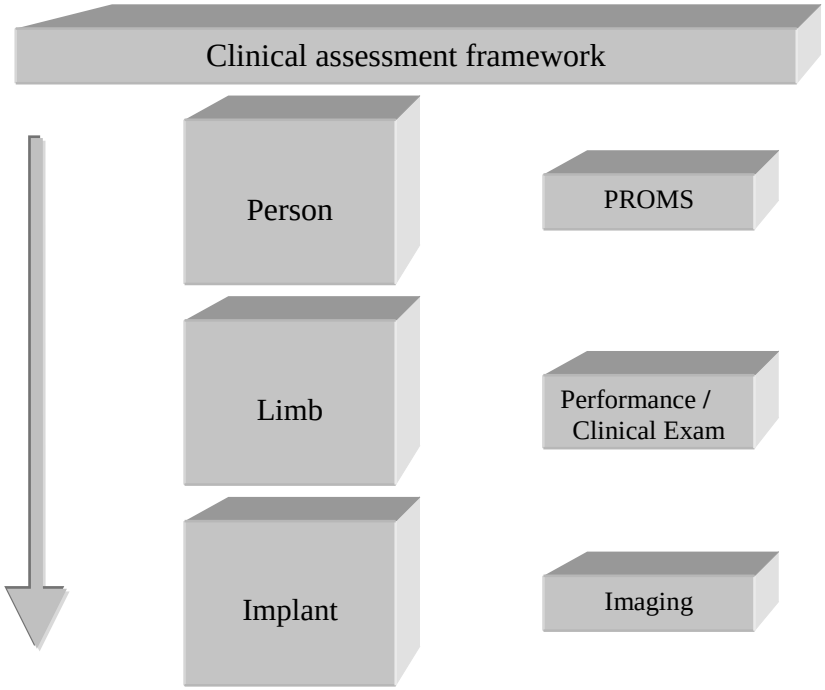
Difference in patient report and direct assessment of outcome is important, as it has been suggested that patient reported assessment can be used instead of direct assessment (Bellamy et al, 1997). It has previously been suggested that while both are acceptable outcome assessments, they are complementary (not equivalent), and should be reported as such (Hamilton et al. conference abstracts, Appendix B). It was previously reported in a study of 100 patients that patient reported outcomes correlate well to each other and to pain, but not to specific functional assessments following knee arthroplasty. Oxford Knee Score and the Physical Component Score (PCS) of the SF-12 questionnaire were found to correlate highly ($r = 0.74$, $p = 0.001$) at both six and twelve months post-op. Comparative correlations to the ALF timed score and power output were found to be poor – modest ($r = 0.2 - 0.4$) for both OKS

and SF-12 PCS (Hamilton et al. Orthopaedic Research Society 2010 extended abstract, Appendix B).

A theoretical assessment framework can be drawn that helps explain some of the differences between the assessment types (Figure 4.4). This framework is loosely based on the WHO classification of impairment, disability and handicap. The arthroplasty (in this case TKA) can be assessed at various levels in the body: at the level of the implant, the limb, or in the person as a whole. Different assessments are required to ascertain information about each level. Performance data generally assesses function of the limb, while patient reported information reflects the issues (related to their pathology, pain, further symptoms or surgery) perceived by the individual about their functioning in society.

It is notable that the strong association between PROMs and pain report highlighted previously occur at the same level of this model, whereas the weaker associations between performance data and self report are comparing across levels, this may help explain some of the variation in results.

Figure 4.4 - Clinical assessment framework. Figure displays a theoretical hierarchical model that demonstrates the differing levels of information assessed by the alternative types of outcome analysis.



The most effective way to consider this model is to take the example of a poorly functioning TKA. Clinically, information is sought about the functioning of the implant specifically via radiographs. Consideration is then given to component positioning or perhaps loosening. At the same time a clinical examination is performed to ascertain the functional limitations of the limb, and the patient's report of their pain and functional difficulties discussed to determine the most appropriate further treatment options.

This example also highlights the hierarchical nature of the model, where appropriate level of assessment is required to determine problems that relate to that level. A poorly functioning implant (perhaps due to malalignment) may influence the limb, manifested in poor range of motion and walking ability, and further affect the persons ability to perform activities of daily living through this limitation and also

via pain. Whilst the dysfunction can be screened by any level of assessment, only radiographs will ascertain the actual underlying issue. This may reflect the relationship between PROMS and performance data, where physical dysfunction of the quadriceps can be inferred by the PROMS but only confirmed by direct assessment. This may explain why the substantial differences observed in lower limb power in the previous chapter did not result in large differences in patient reported score, and confirms that the specific analysis of the power output employed in the previous chapter was required to assess this specific parameter.

Confounding variables such as mental health issues (depression, anxiety) that are known to relate to the outcome of TKA will have largest influence at the level of the person. The results presented here, and those of Mizner et al (2011) and Stratford and Kennedy (2006) suggest that the role of pain is also most influential at this level of the suggested model.

In conclusion, this was the most comprehensive analysis of a large sample of TKA patients to assess the relationship between physical performance and the patient's report of that performance prior to and following surgery. Further, it is the first analysis to present regression models at all assessment points to reflect how the relationship between these two types of assessment changed over time. The patient's report of their function and direct assessment of that function became more closely associated over time following surgery. Pain was found to be the dominant factor that explained the variation in the OKS, and the improved relationship between performance and report of performance improved as the report of pain diminished. The hierarchical clinical assessment model presented suggests a link between direct functional assessment and patient reported function. This conceptual framework helps explain the link between lower limb power output and overall patient function.

5 Influence of the Musculature on TKR Outcome: 'Regenerative Potential' of the Muscle – a Pilot Study

5.1 Introduction

It has previously been demonstrated that the individuals' lower limb power affects their physical functioning following total knee arthroplasty (Lamb and Frost, 2003). Clinically, it is thought that pre-operative physical function and lower limb power are relevant to the patients' post-operative physical outcome, as correlation of these measures has been demonstrated (Faulkner et al, 2010; Lingard et al, 2004).

Thirty percent (30%) strength deficits compared to aged matched controls have been demonstrated in TKA patients (Silva et al, 2003). Measurements of muscle atrophy has not been able to explain this discrepancy, Mizner et al (2005a) found only a third of strength deficits in the early post operative phase could be accounted for by reduction in muscle mass. Longer term deficits are as yet unexplained. No consideration has been given to the 'regenerative potential' of the musculature to explain post-operative power output following knee replacement.

Muscle satellite cells are undifferentiated myogenic precursors that sit in quiescence below the muscle sarcolemma. In normal muscle, satellite cells respond to regenerative cues such as injury or exercise, by proliferating to form myoblasts, which divide a limited number of times before terminally differentiating and fusing to form multinucleated myotubes (Hawke and Garry, 2001; Morgan and Partridge, 2003; Wagers and Conboy, 2005; Zammit et al, 2006) to provide new myonuclei for the homeostasis, hypertrophy or repair of the muscle fibres (Figure 2.5). Any situation that requires either the repair of existing myofibres or the creation of new myofibres such as the rehabilitation process following TKA relies on the ability of

the muscle satellite cell to generate new myoblasts, as has been previously discussed (Chapter 2).

It has been suggested that the number of satellite cells in an individual is variable and factors such as aging (Renault et al, 2002; Kadi et al, 2004; Verdjik et al, 2007) and level of exercise (Chafri et al 2003; Mackey et al, 2007; Kadi et al, 2010) can affect this number. However human studies to date have involved only small numbers of healthy subjects. The relative number of cells in a typical osteoarthritic or knee arthroplasty population is unknown, and no study has considered the effect of the muscle satellite cell on post operative recovery following knee arthroplasty, or any other orthopaedic procedure.

It was hypothesised that; (1) variation would exist in the number of satellite cells in a cohort of total knee arthroplsty patients, that (2) differing levels of recovery of muscle strength following surgery would be apparent, and that (3) this recovery would be influenced by the underlying number of satellite cells.

5.2 Methods

Patient recruitment

Patients were recruited from the operating list of a single consultant surgeon. Inclusion criteria for participation was that patients were suffering from unilateral osteoarthritis of a severity that was suitable for joint arthroplasty, but were otherwise healthy (i.e., not suffering form any comorbidities that would affect their post operative recovery or subsequent physical performance, such as cardiovascular or neurological disease). Further that the planned procedure was ‘routine’ primary total knee arthroplasty, i.e. the first surgical procedure on the osteoarthritic joint, performed with standard implants, without the addition of any augments. Local

approval was granted for the study by Lothian Research Ethics Committee and NHS Lothian Research and Development Management.

Suitable patients were approached at the time of pre-operative assessment, approximately 2 weeks prior to surgery, and were recruited through informed consent. 18 patients were recruited to the study, although 2 were subsequently withdrawn due to surgical complications. Thus 16 patients completed the study protocol.

Of the withdrawn patients, one acquired a deep infection in the early post-operative phase, the other was found at the time of surgery to have significant synovitis and a subsequent post-operative flare-up of previously undiagnosed rheumatoid arthritis. Both patients were unable to undertake the post-operative power assessments.

Power output assessment

Specific assessment of the recovery of muscle strength in the cohort was assessed by use of a Leg Extensor Power Rig (Queens Medical Centre, Nottingham, NG7 2UH) which has been well validated for use with this population group (Lamb and Frost, 2003; Bassey et al, 1990; Robertson et al, 1998). The Leg Extensor Power Rig has been described previously (Chapter 3). The test procedure followed that suggested by Robertson (1998) as was adhered to and detailed in Chapter 3. Application of force accelerated the flywheel from rest, and output was recorded as both maximal wattage (W) generated and relative power to body weight ratio (%) of a single leg extension. Those not able to complete the test were assigned a score of zero as advocated by Lamb and Frost (2003). After each attempt, the force produced was recorded. The highest recorded output was used for analysis as suggested by Barker and Simpson (2004). Body weight (Kg) was measured prior to leg power testing separately at each time point using a calibrated set of Seca 761 Approved Medical Mechanical Floor Scales (Class IIII).

Baseline power output was assessed at pre-operative assessment, then post-operatively at routine outpatient clinical review at 6 weeks and 26 weeks post surgery in a local clinical testing facility at the Royal Infirmary of Edinburgh.

Muscle sampling

The muscle sample was obtained from distal quadriceps by the same surgeon during the knee arthroplasty via the consultant's routine incision for this procedure. The biopsy technique was standardised and the sample of tissue was harvested approx 5 cm proximal to the superior pole of the patella, dependant on the tissue quality of the individual. Samples were immediately fixed (in the operating theatre) in 10% neutral buffered formalin (4% formaldehyde in phosphate buffered saline) then kept in a fridge for 24 hours. Subsequently the samples were immersed in 70% ethanol before embedding in paraffin wax.

Histology

7 μm sections were cut on a microtome; multiple sections were cut from all samples. Mayer's haematoxylin and eosin (H&E) staining was performed to ensure the muscle tissue quality was satisfactory using a standard local protocol (Appendix E).

Satellite cells were identified by an immunofluorescence microscopy protocol using a primary mouse antibody for Pax7 and goat anti-mouse fluorescently labelled secondary. Counterstaining with DAPI (4', 6-diamidino-2-phenylindole), a fluorescent stain that binds to DNA, established the position of the nuclear material within the cell.

Induction of Pax7 is the mechanism by which pluripotent stem cells within the muscle tissue specify into the myogenic precursor satellite cells (Seale et al, 2000;

Hawke and Garry 2001; Buckingham, 2007). Pax7 has been reported to be present in all adult muscle satellite cells, and is a robust marker of these. As discussed in Chapter 2 there is some debate as to the relative benefits of using antibodies against Pax7 or NCAM. Both markers were trialled, but because Pax7 achieved a far greater staining efficiency, this marker was used for the quantification of satellite cells in the present study.

The wax embedded slides underwent a local standard de-waxing protocol, using xylene and graded ethanol baths (Appendix E), then antigen retrieval, using sodium citrate (Appendix E). Sections were then loaded onto sequenza plates for immunofluorescence staining.

Slides were washed in PBS (without calcium and magnesium) before adding 3 drops Protein Block (AMS biotechnology Abingdon, Oxon, OX14 4RX cat DPB-125) solution for 30 minutes at room temperature. Primary mouse anti-human Pax-7 antibodies were diluted (1:150), to appropriate concentration in the protein block solution and 250µl was added to each slide for 1 hour at room temperature. The slides were washed three times in PBS, and Alexa Fluor (Invitrogen) goat anti-mouse 488 (red) 1:500 secondary antibodies were added for 30 minutes at room temperature in the dark. A further 4 washes in PBS were done before dehydrating the slides in 100% ethanol and drying. The slides were then mounted with Vectashield hardest mounting medium with DAPI for fluorescence (Vector laboratories, Burlingame, CA 94010, cat H-1500).

A control experiment was run in parallel with all sections that followed the exact protocol, but did not include primary antibodies for each section to ensure antibody specific staining.

Images of the sections were captured on a Zeiss Axioscope II MOT compound microscope under a 20x objective equipped with a Hamamatsu Camera Controller

(ORCA-ER) and OpenLab 4.0 image analysis software (ImproVision, Coventry, UK). Fluorescent images were then merged using Photoshop (Adobe®). Positive identification of satellite cells was determined by immunofluorescent staining in association with nuclear material, and by position under the basal lamina.

Three separate areas of each section were imaged and counted separately. In each image, the number of muscle fibres, satellite cells and myonuclei were counted, the average count of the section was used for analysis. The counting of cells on each image was repeated three times to ensure consistency. The cell positive staining index (PSI) was calculated (no of satellite cells/total no of myonuclei x 100). The count was made by a single individual (the author) blinded to the patient identity until the analysis was complete.

Statistical analysis

Statistical analysis was carried out with the Minitab (release 15) software. All data was assessed visually for normality. Differences in power outcome scores across the 3 assessment time points were assessed by paired sample t-tests. Strength of correlation between power output and satellite cell number was assessed with the Pearson product moment correlation coefficient and simple regression analysis performed to quantify this association. Multiple linear regression analysis was further employed using a stepwise model building technique screen out predictors not associated with the response. Level of significance was accepted as $p = 0.05$.

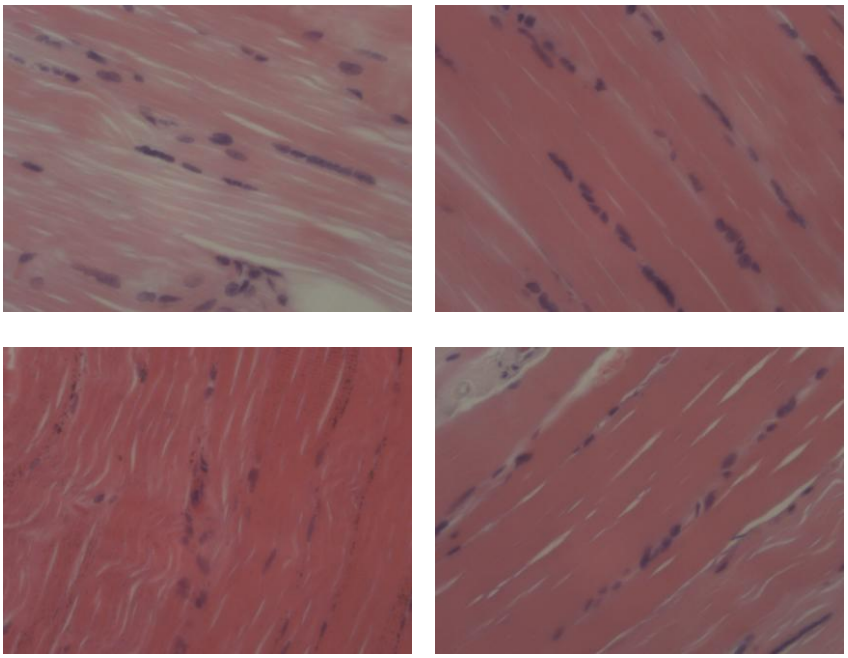
5.3 Results

Histology

One sample was unfortunately damaged in the dewaxing process and all further staining attempts were unsuccessful. All of the remaining 17 samples were found to

contain viable muscle tissue (Figure 5.2). The control experiment, using only secondary antibodies, resulted in no staining thus demonstrating that there was no non-specific staining.

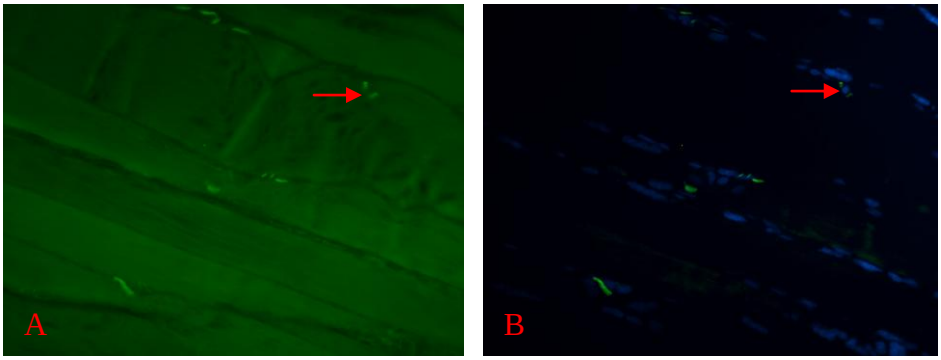
Figure 5.1 - Examples of typical H&E staining of study tissue sections. Normal muscle architecture demonstrated in all samples. Haematoxylin stains the nuclei blue and eosin the collagen fibres pink.



Positive cells were identified on the basis of positive immunofluorescent staining in association with nuclear material, and confirmed by their position under the basal lamina. In cases where the final image confirmed the presence of Pax-7 and DAPI, but the location under the sarcolemma was in question, the specific immunofluorescence image for Pax-7 (i.e. not combined with the image highlighting the DAPI stain) was viewed. On this image the muscle architecture was clearer; positive cells superficial to the basal lamina were not counted. Analyses of these two images enable classification. An example is portrayed in Figure 5.2, where the final image (B) suggested the presence of a satellite cell towards the upper right corner of the image (green fluorescence around the blue nuclear stain), but the location was not consistent with the overall architecture of the muscle myonuclei. Analysis of the

'green channel' revealed the position of this cell to be superficial to the basal lamina. This was not therefore counted as a satellite cell.

Figure 5.2 - Multiple images of the section to confirm cell location. Immunofluorescent stain, merged with DAPI stain of myonuclei

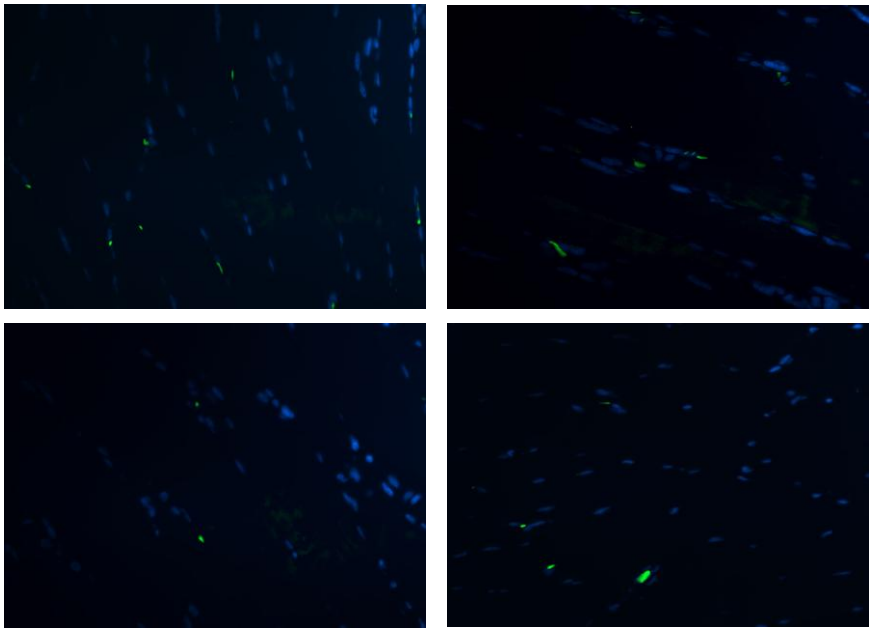


A = 'green' Pax-7 immunofluorescence stain, B = merged and adjusted image (Pax-7 and DAPI stains)

Numbers of identified satellite cells

Variation of satellite cell number was observed within the patient population (Figure 5.3 and Table 5.1). The positive staining index (PSI) of muscle satellite cells was calculated by assessing three separate areas of each individual section. Mean values per individual patient/section are reported in Table 5.1. The satellite cell data was normally distributed, with a mean PSI of 7.62 (95% confidence interval of mean 6.30, 8.94) satellite cells per sample. Variation was found in the cell numbers (PSI 3.07 – 11.35) per individual patient (Table 5.1).

Figure 5.3 – Immunofluorescence image of muscle satellite cells. Examples of typical staining profiles obtained. Separate samples demonstrating differing numbers of satellite cells identified with the staining protocol.



Satellite cells are identified with an antibody against Pax7 (green), further myonuclei with DAPI (blue).

Table 5.1 - Satellite cell count

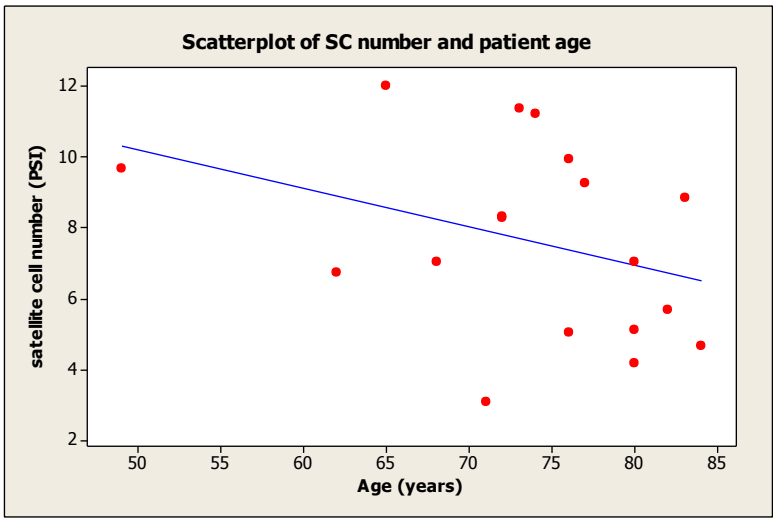
Sample	Cell count	Myonuclei	Fibres	PSI
2	5.34	105	23	5.09
3	13	108.34	16.67	11.99
4	9.67	104.67	17.67	9.24
5	8.4	101.67	16.34	8.26
6	12.25	123.34	11.5	9.93
7	9.25	104.5	20.25	8.85
8	8	82.67	14	9.68
9	8.67	104.34	15	8.31
10	4	130.34	18.25	3.07
11	10.67	94	44.67	11.35
12	7.34	104.34	16.67	7.04
13	5	107.67	28.67	4.64
14	11	98.25	21.75	11.20
15	5.67	136.24	24.34	4.16
16	5.67	100.34	38.67	5.65
17	9.5	135	14.5	7.04
18	10.67	158.67	43.34	6.73

Average count of 3 imaged regions per muscle section/individual patient. Samples re-coded for blinding purposes, Sample 1 represents the destroyed sample and is not displayed.

Correlation of satellite cell number with patient age

Patient age in the cohort was normally distributed, with a mean age of 73.5 (95% CI of mean 69.3, 77.9) years. The variation in satellite cell number formed a modest linear association with the age of the patient (r = -0.37, p = 0.14), see Figure 5.4, where the biopsies of the younger patients tended to contain greater numbers of satellite cells than the older patients.

Figure 5.4 - Association of patient age and satellite cell number. Linear relationship found between patient age at time of surgery and satellite cell PSI.



Power output

All data was normally distributed. Wide variation in power output between individual patients was observed, see table 5.2. Mean maximal power output of the cohort was 48.2 (SE of mean 12.6) W at pre-op, 60.2 (9.98) W at 6 weeks post-op and 82.4 (11.2) W at assessment 26 weeks post-op. This reflected a power output to body weight ratio of 53.7 (14.3) % at pre-op assessment, 75.0 (12) % at 6 week assessment and 101.9 (12.3) % at 26 week assessment.

The large standard errors of the mean reported reflect the large individual variation in maximal power output amongst the patient group.

Table 5.2 – Mean individual power output at assessments

Sample	Output at assessment						Change between assessments			
	Pre-op		6 weeks		26 weeks		Pre-op - 6 week		6 - 26 weeks	
	W	%BW	W	%BW	W	%BW	W	%BW	W	%BW
1	30	0.4	33	0.5	41	0.6	3	0.1	8	0.1
2	18	0.2	72	1	74	1	54	0.8	2	0
3	116	1.1	104	1	134	1.2	-12	-0.1	18	0.2
4	33	0.3	69	0.7	103	1	36	0.4	24	0.3
5	73	0.8	60	0.8	103	1.3	-13	0	43	0.5
6	1	0	35	0.4	55	1	34	0.4	20	0.6
7	2	0	8	0.1	36	0.4	6	0.1	28	0.3
8	Data unavailable									
9	137	1.8	148	2	189	2.4	11	0.2	41	0.4
10	97	0.6	86	0.8	84	0.7	-9	0.2	-2	-0.1
11	59	1.1	46	0.7	65	1.2	-13	-0.4	20	0.5
12	38	0.5	70	1.1	72	1.1	32	0.6	2	0
13	4	0.1	20	0.3	28	0.4	16	0	8	0.1
14	13	0.2	65	0.9	92	1.2	52	0.2	33	0.3
15	3	0	1	0	38	0.6	-2	0	11	0.3
16	Data unavailable									
17	144	1.5	115	1.2	149	1.5	-29	-0.3	34	0.3
18	3	0	31	0.5	56	0.7	28	0.5	25	0.2

W = maximal power output, measured in watts, %BW = maximal power output expressed as power to patient body weight ratio.

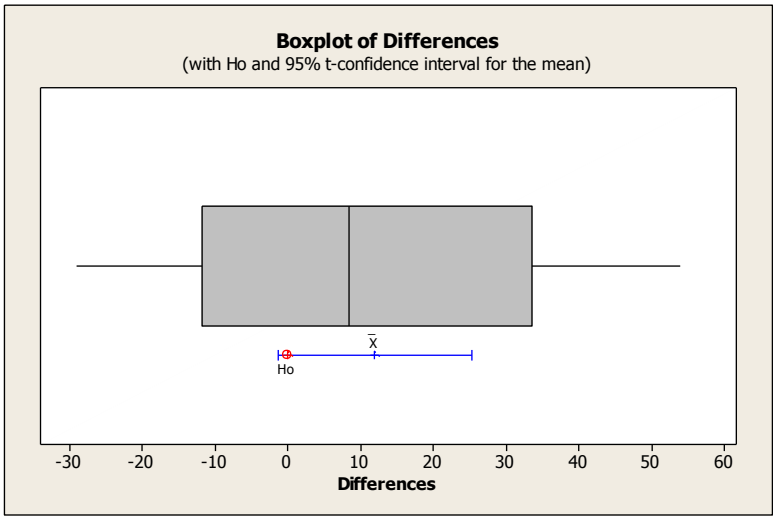
Change in power output between assessments

Substantial difference between individuals was noted. Improvement of maximal power output varied between 0 and 70 W, which reflected a change of -30% – 80% in power to body weight ratio.

Improvement in power output was generally observed between assessment points. Mean improvement was 12.12 W between pre-op and 6 week assessment, though this difference was not statistically significant, paired samples t-test, 95% CI for mean difference: -1.37, 25.37 $p = 0.075$. Again, the wide variation in individual

results accounted for the large confidence intervals of the mean difference. Figure 5.5 shows a boxplot of the difference in mean power output between pre-op and 6 week assessments. This demonstrates the wide spread of data, and the resultant wide 95% confidence intervals, that narrowly incorporate the zero value.

Figure 5.5 - Improvement in power output between pre-operative and 6 weeks values



Ho = null hypothesis of no difference existing between mean values at pre-op and 6 week assessment, x = estimate of the difference in power output between the pre-op and 6 week assessment periods

Mean improvement in maximal power output between 6 and 26 weeks was 19.68 W. This difference was statistically significant, 95% CI for mean difference: 14.43, 30.07 p = <0.000.

The improvement in maximal power output reflected an improvement of power-body weight ratio of 21% between pre-op and 6 weeks. This difference was statistically significant 95% CI for mean difference: 0.03, 0.39 p = 0.025. A further 27% mean improvement in power to body weight ratio was found between 6 and 26 week assessments, which was also statistically significant 95% CI for mean difference: 0.15, 0.38 p = <0.000.

Correlation of satellite cell content and power output

Good correlation was noted between the patients satellite cell number and improvement in power output between 6 weeks and 26 weeks, both in maximal power output ($r = 0.57$, $p = 0.023$) and power-body weight ratio ($r = 0.64$, $p = 0.008$) (Table 5.3, Figures 5.6 and 5.7).

Table 5.3 - Correlation of satellite cell number and power output

Power Output					
Pre-op		6 weeks		26 weeks	
Power	%BW	Power	%BW	Power	%BW
r = 0.13 p = 0.62	r = -0.26 p = 0.33	r = 0.21 p = 0.44	r = 0.21 p = 0.44	r = 0.34 p = 0.2	r = 0.41 p = 0.12
Change in Power Output					
Pre-op – 6 wks		6 – 26 weeks		Pre-op – 26weeks	
Power	%BW	Power	%BW	Power	%BW
r = 0.06 p =0.83	r = -0.23 p = 0.36	r = 0.57 p = 0.02	r = 0.64 p = 0.01	r = 0.32 p = 0.23	r = 0.16 p = 0.55

Correlation coefficients of cell number and lower limb power output at the various time points for both power at that assessment, and change in power between assessments. Significant correlations highlighted in red.

Figure 5.6 - Linear relationship found between positive staining index of satellite cells in muscle tissue samples and the subsequent change in individual maximal power output. Good correlation between variables r = 0.57, p = 0.02

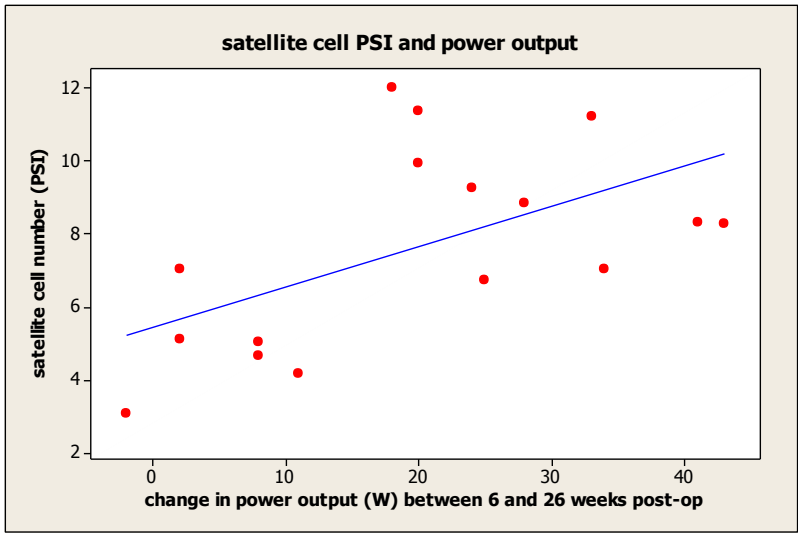
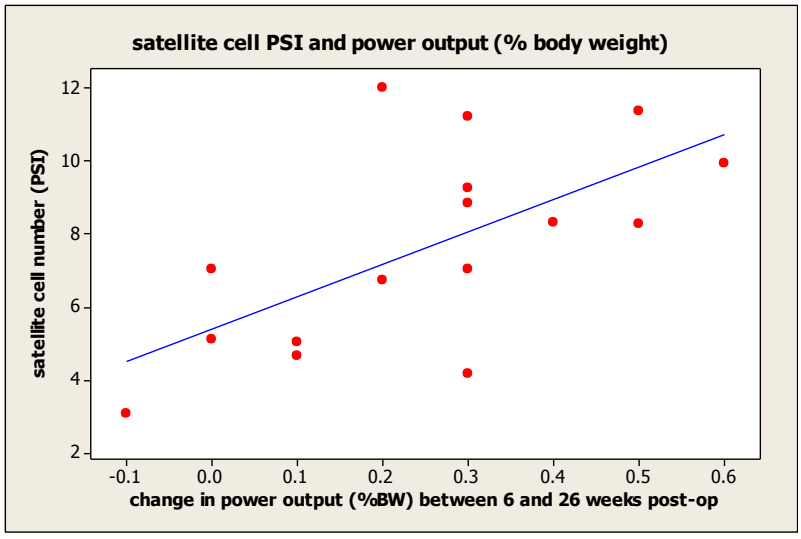


Figure 5.7 - Linear relationship found between satellite cell positive staining index and individual change in power-bodyweight ratio, strong correlation between the variables r = 0.64, p = 0.008



Regression analysis

Uni-variant regression analysis demonstrated that 27 % of improved maximal power output between 6 and 26 weeks could be attributed to the variation in the number of underlying satellite cells (R^2 adjusted = 27.0%). Further, 37% of the variation in the improvement of power to body weight ratio could be attributed to the satellite cell pool (R^2 adjusted = 36.6 %). Full analysis is displayed in Appendix F.

Poor correlation was found between both pre-op power output and improvement in power output (r = 0.32, p = 0.23) and pre-op power-body weight ratio and improvement in power-bodyweight ratio (r = 0.19, p = 0.47). Uni-variant regression demonstrated that pre-op scores account for very small amounts of the change in maximal power output post-op (R^2 adjusted = 3.5%) and were not indicative of improved power-bodyweight ratio. Full analysis is displayed in Appendix F.

Stepwise regression modelling was performed (alpha to enter or remove p = 0.05) to assess the relative contribution of each factor to the change in power output. The only factor associated with change in post-operative power output (both maximal watts and proportion of body-weight) was the number of satellite cells found at time of surgery (see Appendix F for statistical data output).

5.4 Discussion

This is the first study to assess the influence of the underlying regenerative potential of the muscle tissue of the extensor mechanism group on post-operative power output following total knee arthroplasty. A third of the variation of change in power output could be attributed to the intrinsic number of muscle satellite cells in the quadriceps muscle at the time of surgery. This result was substantial, and can put into context by the associated result that only 3.5% of the variation in change in post-

operative power could be accounted for by the commonly used clinical measure of pre-operative power score. Further regression modelling in this cohort found that, of the two factors, only the satellite cell number was predictive of the post-operative change in power output (Appendix F).

As described in Chapter 2, the muscle satellite cell is responsible for the generation of myoblasts that fuse to repair or to form new muscle fibres (Figure 2.5). The rehabilitative process following TKA typically results in training effects and hypertrophy of the quadriceps muscle that influence post-operative outcome (Greene and Schurman, 2008; Mizner et al 2005a). This process is driven by the differentiation of satellite cells through the myogenic program as previously outlined.

.

Pain (attributable to OA) is the prime indication for undergoing knee arthroplasty, and reduction of this pain the major goal of surgery. It is likely that pre-operative pain prevents accurate measurement of power output at this time, by limiting the patients' ability to generate maximal force through a pain inhibition process. In keeping with this no correlation was observed with the satellite cell number and the change of pre-operative power to either 6 or 26 weeks. Much of the surgical pain had abated by the 6 weeks review, and the individuals were all able to mobilise without the use of aids that would have been necessary in the early post-operative phase.

Improvement in muscle power output has been demonstrated by 6 months post TKA (Lamb and Frost, 2003). This current study corroborated that report demonstrating an improved mean muscle power between assessments. A mean difference of 12W was found between pre-op and 6 week review, then a further mean difference of 20W between 6 and 26 weeks. Apart from the underlying osteoarthritic knee, the patients were otherwise healthy, thus no other comorbidies were present which would have acted as confounding factors for the physical recovery post-op.

Magnitude of power output is inherently associated with the individual's underlying muscle mass, as every muscle fibre can be thought of as an independent force generating unit (Jones and Round, 1990). For this reason, measurement of the improvement in power was assessed both in terms of absolute power and ratio of power relative to bodyweight. The power to bodyweight figure takes into account the general size of the individual, and is therefore the more relevant of the two measures in assessing change expressed in the population. It has been demonstrated that muscle power correlates more strongly than muscle strength to physical function, as power involves both force production and contraction velocity, which better reflects daily tasks that are force and speed dependant (Evans, 2000, Bassey and Short 1990, Bean et al, 2003). Interestingly, of the results presented here it is the power-bodyweight ratio that demonstrates the strongest correlation with the satellite cell count ($r = 0.64$, Figure 5.7). The number of satellite cells also accounts for the largest percentage of change in power output when assessed as power-weight ratio (R^2 = 36.6%).

Mizner et al (2005a) reported a reduction of quadriceps strength in the first 3-4 weeks post-operatively compared to pre-operative levels in a small knee arthroplasty cohort. They demonstrated that failure of voluntary muscle activation and muscle atrophy explained 85% of the loss of quadriceps strength ($p < 0.001$). Multiple linear regression analysis further revealed that failure of voluntary activation contributed around twice as much as atrophy. This failure of neuromuscular activation in the early post-operative phase is an attractive explanation of the poor levels of quadriceps strength and power reported in the first month post-operatively, however does not well explain the reported prolonged power and strength deficits compared to healthy controls (Silva et al, 2003; Huang et al, 1996).

No correlation was found in this study between satellite cell number and power output at the 6 weeks assessment (broadly around the same time point as the Mizner study). It is possible that the neuromuscular function deficits are important in the immediate post-operative period, whereas improvement in power output beyond this

early stage may be more closely associated with regenerative potential and satellite cell content. The difference in detected power levels between this study and those reported by the Misner group suggest either that the majority of reduction of the surgical pain that follows knee arthroplasty occurs between 4 and 6 weeks, or that the influence of a failure of voluntary activation is moderated by the 6 weeks assessment.

It is notable that the previously reported decrease in cell number with advancing age was detected despite the relatively limited age range typical of a primary knee arthroplasty population. The modest strength of the correlation is indicative of this limited age range, and may be strengthened with increased data. This is also the first study to report variation in the individuals' satellite cell numbers in an osteoarthritic population. This variation was expected and corroborates those reports of the wider population (Kadi et al, 2004; Renault et al, 2002; Verdjik et al, 2007). The process of osteoarthritis is not known to affect the number of individual satellite cells; however patients with osteoarthritis are likely to exercise less. Chafri et al (2003) have demonstrated a preservation of the satellite cell pool in elderly subjects that engaged in regular exercise. This may suggest that a typical osteoarthritic group may present with a comparatively reduced cell number due to the poor levels of activity generally associated with the disease. Individual variation depended on activity levels, and this was found in this cohort.

The detection of Pax7 in the muscle section is indicative of satellite cells, as Pax7 is specific to muscle satellite cells. It is present in all adult myogenic precursor cells due to its role in specifying the differentiation of pluripotent stem cells (Seale, 2000). Previous human samples in healthy populations (Kadi et al, 2006, Chafri et al, 2003, Lindstrom and Thornell, 2009) have successfully used NCAM as an additional satellite cell marker. Staining with an NCAM marker was also attempted in this study, though the Pax7 marker was used to quantify the satellite cells due to a far greater staining efficiency. NCAM is a cell surface marker, and it may be that the embedding and retrieval from paraffin wax caused damage that the cytoplasmic Pax7

marker was not subjected to. It is notable that the other human studies (Kadi et al, 2004 and 2006, Chafri et al, 2003, Lindstrom and Thornell, 2009) used frozen samples with this cell marker which would not have been subject to this procedural issue. The specific positive identification criteria of further association with the DAPI stained nuclear material, and also physical location under the basal lamina helped reduce the possibilities of false positive detection.

In conclusion, pre-operative lower limb power (and thus by correlation function) was found to be less predictive of physical muscle power recovery post TKA than the number of satellite cells within the quadriceps muscle. Ten times the variation of change in power output post operatively was linked to the underlying number of muscle satellite cells in the individual compared to a measure of pre-operative power output, and stepwise regression modelling found only the number of cells to be an associated variable.

6 Influence of the Musculature on TKR Outcome: Satellite Cell Activation and Power Output

6.1 Introduction

The results presented in the previous chapter suggested that the inherent number of satellite cells in the quadriceps muscle may substantially influence the patient's post-operative lower limb power output, and thus function, following knee arthroplasty. These results were obtained on thin longitudinal sections of the sample tissue which may not be representative of the content of the whole muscle.

To explore this further real time reverse transcription polymerase chain reaction (RT-PCR) analysis was used as the whole biopsy sample could be assessed to determine the quantity of specific genetic material. It enables both detection and quantification (as a relative amount) of specific gene expression in a sample.

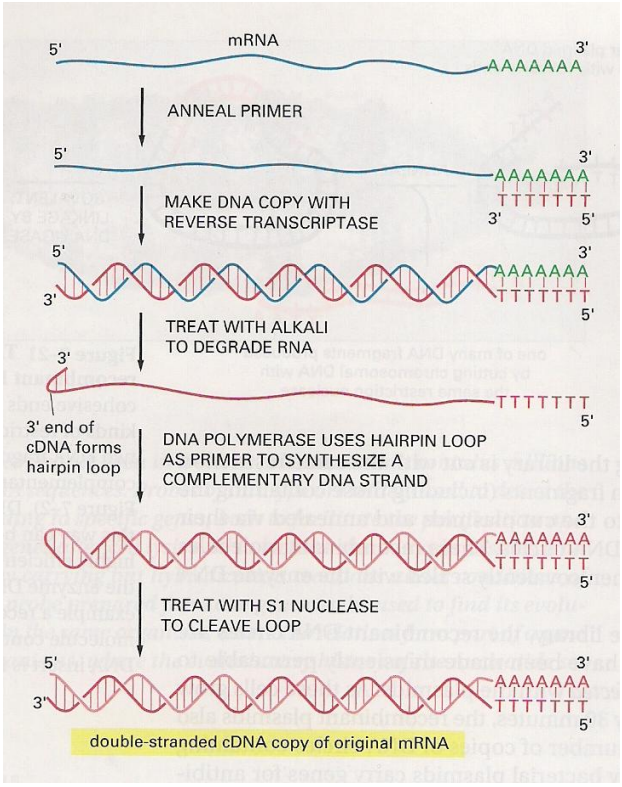
The PCR technique was devised in 1985 and has revolutionised molecular biology, as the recombinant DNA methods previously employed took weeks to clone a segment of DNA, whereas with PCR billions of copies can be made in a few hours. A further benefit to previous methods is the use of primers to determine the specific gene sequence to be amplified, thus isolation of the desired segment from the whole sample is not required (Campbell, 1996).

The technique uses a solution of double-stranded DNA containing the nucleotide sequence that is to be targeted for copying. The enzyme DNA polymerase is added as a catalyst, along with a supply of all four nucleotides (to assemble the new DNA) and specific primers. These primers are chemically synthesised with sequences that are complimentary to the ends of the targeted segment of DNA, and are required for the DNA polymerase enzyme to initiate the DNA synthesis.

The actual process involves heating the DNA to separate the strands, then cooling to allow the primers to attach to (via hydrogen bonding) to the ends of the target sequence. The DNA polymerase extends the primers by adding nucleotides, using the longer DNA strands as a template. Further heating begins the next cycle of strand separation, primer binding and DNA synthesis (Campbell, 1996).

Double stranded DNA is required for PCR, though it is single stranded specific mRNA that is extracted form the tissue samples. Normal transcription in the cell nucleus involves the synthesis of mRNA from DNA, the mRNA going on to synthesize proteins in the cell cytoplasm. Reverse transcription is the reverse of this process where a DNA copy (cDNA) is produced by the enzyme reverse transcriptase (a viral enzyme that uses an RNA strand as a template for the synthesis of a complementary DNA strand) forming a cDNA/mRNA hybrid helix that can be used for PCR (Albertz et al, 1994).

Figure 6.1 - Diagram detailing the process of reverse transcription, where cDNA is produced from mRNA using the enzyme reverse transcriptase.

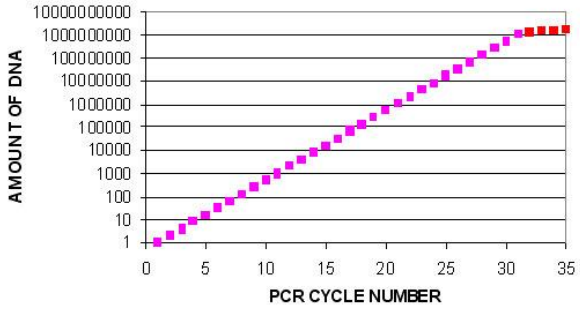


Taken from: Albertz B, Bery D, Lewis J, Raff M, Roberts K, Watson J. Molecular Biology of The Cell, Fourth Edition, 1994, Garland Publishing Inc, New York.

Measurement of DNA amplification during PCR is in real time, i.e., the amplified product is measured during each PCR cycle. As PCR amplification is an exponential reaction, a straight line relationship between the amount of DNA and cycle number is apparent when plotted on a logarithmic scale (Figure 6.2 A). In the standard output plot of PCR data (Figure 6.2 B), the early doublings are not apparent at the scale used. A threshold is set at which the sample expresses beyond any background levels, and the cycle at which the sample crosses the threshold is called the cycle threshold, C_t. The quantity of DNA theoretically doubles every cycle during the exponential phase (Figure 6.2 A) and relative amounts of DNA can be calculated.

Figure 6.2 – Explanation of the typical PCR output graph demonstrating the typical PCR output graph and logarithmic relationship.

A



B

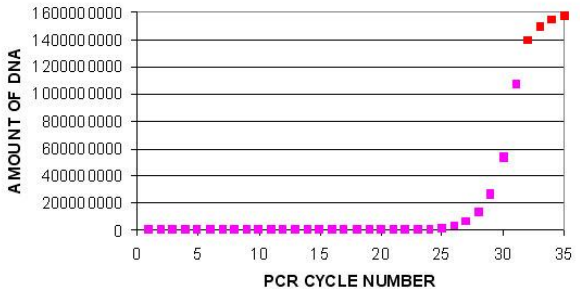


Figure adapted from Applied Biosystems promotional literature

Measurement of DNA is in relation to a housekeeping gene measured in the same sample to normalise for possible variation in the amount of DNA between samples.

This normalization permits accurate comparison of expression of the gene of interest between different samples, provided that the expression of the housekeeping gene used in the normalization is very similar across all the samples (Nailis et al, 2006; Nolan et al, 2006).

The use of this technique also allows for the investigation of the expression of genes relevant to different stages of the cell cycle. It is hypothesised that the number of activated satellite cells should correlate more strongly with power output than the number of quiescent cells.

Secondary research questions were to: (a) corroborate the results of the previous study, using immunofluorescence staining protocols, and explore variation in the individual expression of Pax-7 and thereby satellite cell content. (b) correlate muscle satellite cell content to the ability of the patient to improve power output post knee arthroplasty, as was previously demonstrated.

6.2 Methods

Patient recruitment and power output analysis

Patient recruitment was from the operating list of the same surgeon, as per the preceding study. Inclusion criteria were identical, briefly, that the patients were undergoing a planned primary total knee replacement for osteoarthritis and were free of co-morbidities that would affect their subsequent ability to generate lower limb power. Local approval was granted for the study by Lothian Research Ethics Committee and NHS Lothian Research and Development Management.

Suitable patients were approached at the time of pre-operative assessment, approximately 2 weeks prior to surgery, and were recruited through informed consent. 11 patients were recruited to the study.

Specific assessment of the recovery of muscle strength in our cohort was assessed by Leg Extensor Power-Rig (in the manner previously described) and output reported both by maximal wattage (W) generated, and in relative power to bodyweight ratio. Baseline power output was assessed at pre-operative assessment, then post-operatively at routine outpatient clinical review at 6 weeks and 26 weeks post surgery in a local clinical testing facility at the Royal Infirmary of Edinburgh.

Muscle sampling

The muscle sample was obtained from distal quadriceps by the same surgeon during the knee arthroplasty via the standard incision for this procedure. Sampling was standardised, as per the previous experiment, aiming to biopsy the quadriceps 5 cm superior to the patella, dependant on the tissue quality of the individual. Samples were immediately snap frozen in liquid nitrogen (in the anti-room of the operating theatre) then subsequently transferred to -80°C storage.

A control biopsy of quadriceps muscle (required for the RT-PCR analysis) was obtained from a young patient undergoing limb length correctional surgery. Attempts were made by the surgeon to biopsy from the quadriceps group as close as possible to the biopsy site in the knee arthroplasty group. This was also immediately snap frozen in liquid nitrogen and processed as per the arthroplasty samples. The sample was obtained via informed consent.

Histology

Frozen samples were transferred to -20°C for 24 hours before being sectioned. Four 7μm thick tissue sections per sample were cut on a Leica 1850 cryotome for immunofluorescence analysis. The remainder of the samples were immersed in RNAlater ice (Ambion) to preserve the RNA and kept at -80°C prior to for RT-PCR analysis.

Immunofluorescence for Pax7

An antigen against the satellite cell marker Pax7 was use as per the previous experiment. Frozen sections were fixed in cold acetone for 5 minutes and left to air dry before loading onto sequenza plates. Slides were washed in PBS (w/o) before

adding 3 drops of Protein Block (AMS biotechnology Europe ltd) solution for 30 minutes at room temperature. Primary antibodies were diluted, to appropriate concentration in the protein block solution and 250µl was added to each slide for 1 hour at room temperature. The slides were washed three times in PBS and Alexa Fluor (Invitrogen) secondary antibodies added for 30 minutes at room temperature in the dark. A further 4 washes in PBS were completed before adding 250µl of 1µg/ml DAPI, which was followed by 2 washes with water. The slides were then dehydrated in 100% ethanol and dried before the slides were mounted with coverslips in Vectashield Hard Set mounting medium for fluorescence (Vector).

A control experiment was run in parallel with all sections that followed the exact protocol, but did not include primary antibodies for each section to ensure antibody specific staining.

RNA extraction

Prior to RNA extraction, the samples were thawed at -20 degrees for 24 hrs. Tissue was chopped finely with a scalpel blade, and then RNA extraction was performed using nucleospin columns (Macheray Nagel) which bind to the nucleic acid (specific protocol Appendix G). Quantification of the RNA yielded was performed on a nanodrop spectrophotometer. The amount of mRNA detected per sample is expressed in Table 6.1.

Table – 6.1 - RNA quantity in frozen samples

Sample Number	RNA (µg)	Amount of RNA (µl) needed to make 0.3µg of cDNA	Additional H_20 (µl)
Control	0.044	6.8	3.2
1	0.003	10	0
2	0.086	3.5	6.5
3	0.056	5.4	4.6
4	0.045	6.7	3.3
5	0.072	4.2	5.8
6	0.003	10	0
7	0.088	3.4	6.6
8	0.11	2.7	7.3
9	0.038	7.9	2.1
10	0.057	5.3	4.7
11	0.049	6.1	3.9

The cDNA reverse transcription kit requires a 10µl volume of mRNA containing equal amounts of mRNA per sample. The lowest amount of mRNA detected sets the amount per sample. In this experiment, sample 7 contained the lowest concentration of mRNA (0.3µg), thus the other samples were diluted to this concentration (see table 6.1) yielding a solution of 10µl of mRNA at concentration of 0.3µg per µl. the reverse transcriptase kit (AB) adds a further 10 µl of a master mix containing dNTPs, buffer, random primers, RNase inhibitor, reverse transcriptase and water (Appendix G, high capacity cDNA reverse transcription protocol). cDNA is thus made at a concentration of 0.3µg per µl which is then used for RT-PCR analysis.

Quantitative Real Time – PCR analysis

The SYBR green method of quantification was used. SYBR green is a fluorescent dye that binds double-stranded DNA, and upon excitation emits light. Thus, as a PCR product accumulates, fluorescence increases. The qPCR reads the level of SYBR green bound to each cycle. The cycle number SYBR green is detected above the background threshold determines how abundant the gene is. This is compared with a house keeping gene to determine how much of the gene in question is contained in the sample.

Primers

Specific intron spanning primers were designed for the satellite cell markers Pax7, NCAM and CD-34 by technical staff in the ELIGI lab at the Queens Medical Research Institute, Edinburgh University and made externally by Eurogentec.

Satellite cell markers

Pax7 is considered to be a robust marker for satellite cells as discussed previously (Seale et al, 2000), NCAM is a marker of activated satellite cells (Lindstrom and Thornell, 2009), and CD34 (a cell surface glycoprotein which functions as a cell-cell adhesion factor) which has been proposed as a marker of quiescent cells (Beauchamp et al, 2000).

The commonly used 18s (ribosomal RNA) primer was included as the housekeeping gene, (RT-PCR 18S control kit Eurogentec).

Quantitative PCR

To quantify gene expression in human muscle tissue, each reaction contained 1μg cDNA from each patient/sample. Per sample (each well of 96 well plate) 10ul of Syber Green (iTaq SYBR Green Supermix with rox, BioRad) was added along with 7ul of water, 1ul of forward primer, 1ul of reverse primer (Eurogentech) and run in duplicate on a 7500 Fast Real-time PCR machine (Applied Biosystems). The following PCR cycling conditions were used for amplification: 50deg 2 min, 95deg 10 min, then 40 cycles of 95deg 15 sec and 60deg 1 min. Dissociation step of 95 deg for 15 sec, 60deg for 1 min and 95deg for 15 sec were performed for primer specificity, to reduce the chance of false positive results.

Enough volume of cDNA was extracted from the mRNA in the samples to allow two separate runs of PCR. All four markers were included in the two individual plates (in duplicate) for all 12 samples. Average values were calculates for the plates, and then for the two runs for internal consistency.

Statistical analysis

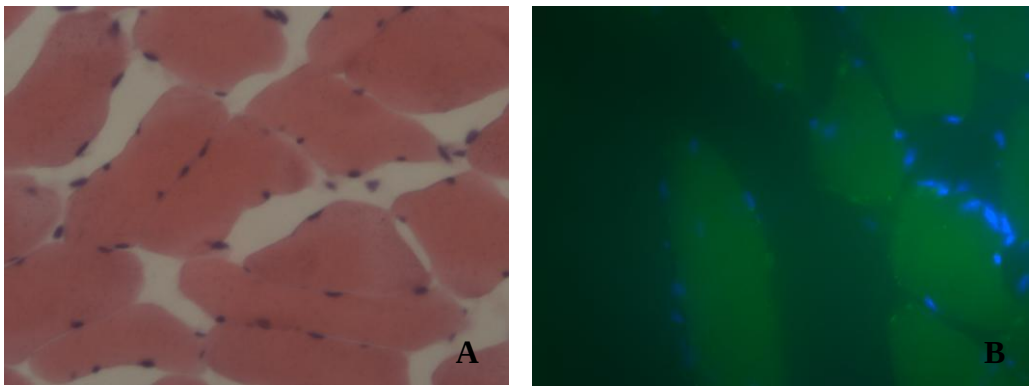
Statistical analysis was carried out with the Minitab (release 15) software. All data was assessed visually for normality. Differences in power outcome scores across the three assessment time points were assessed by paired sample t-tests. Strength of correlation between power output and satellite cell number was assessed with the Pearson product moment correlation coefficient and regression analysis performed to quantify this association. Multiple linear regression analysis was further employed using a stepwise model building technique screen out predictors not associated with the response. Level of significance was accepted as $p = 0.05$.

6.3 Results

Immunofluorescence

Unfortunately it was retrospectively determined that temperature regulator of the cryotome was malfunctioning at the time of sectioning. This has caused the samples to thaw during the cutting process, which has resulted in poor sample integrity (Figure 6.3). All sections were cut from the samples at the same time, thus all slides, and control slides were similarly affected.

Figure 6.3 - H &E and Immunofluorescent staining, highlighting destruction of muscle architecture. Examples of typical H&E staining (A) and immunofluorescence staining (B) of study tissue sections. Abnormal muscle architecture with disruption of myofibres demonstrated in all samples. Image A: Haematoxylin stains the nuclei blue and eosin the collagen fibres pink. Image B: DAPI nuclear staining (Blue) and Pax-7 fluorescent stain (bright green). It was not possible to confirm the location of the stained cells relative to the basal lamina.



Images taken with a Zeiss Axioscope II MOT compound microscope under a 20x objective equipped with a Hamamatsu Camera Controller (ORCA-ER).

The immunofluorescent staining protocol stains the tissue satisfactorily although the muscle architecture has been destroyed as a result of the thawing at time of sectioning. An accurate count of the satellite cells that satisfies the previous criteria is not possible as location under the basal lamina could not be confirmed.

RT-PCR

Ct values were calculated for the expression of the three genes within each sample (using 18S as a reference gene, see figure X). ΔCt values (difference between cycle number of housekeeping gene and gene in question) were calculated for the three satellite cell markers (Figure 6.4 and Table 6.2).

As expected the expression of Pax7 (generic marker) was detected before the NCAM (activated cell marker) suggesting a larger volume of this in the samples. Detecting

the Pax7 expression three full PCR cycles prior to the NCAM expression indicated that there was eight times as much Pax7 as there was NCAM in the samples. The CD34 marker of the inactive cells was expressed approximately two cycles ahead of the Pax7 suggesting larger volumes of this marker than the definitive Pax7 content.

Figure 6.4 - Screen shot of PCR output graph showing 18s and Pax7 expression of all 11 samples. ΔCt values (difference between cycle number that expression occurs) of 18s and Pax7 are highlighted by the black arrow bars for one of the 11 samples.

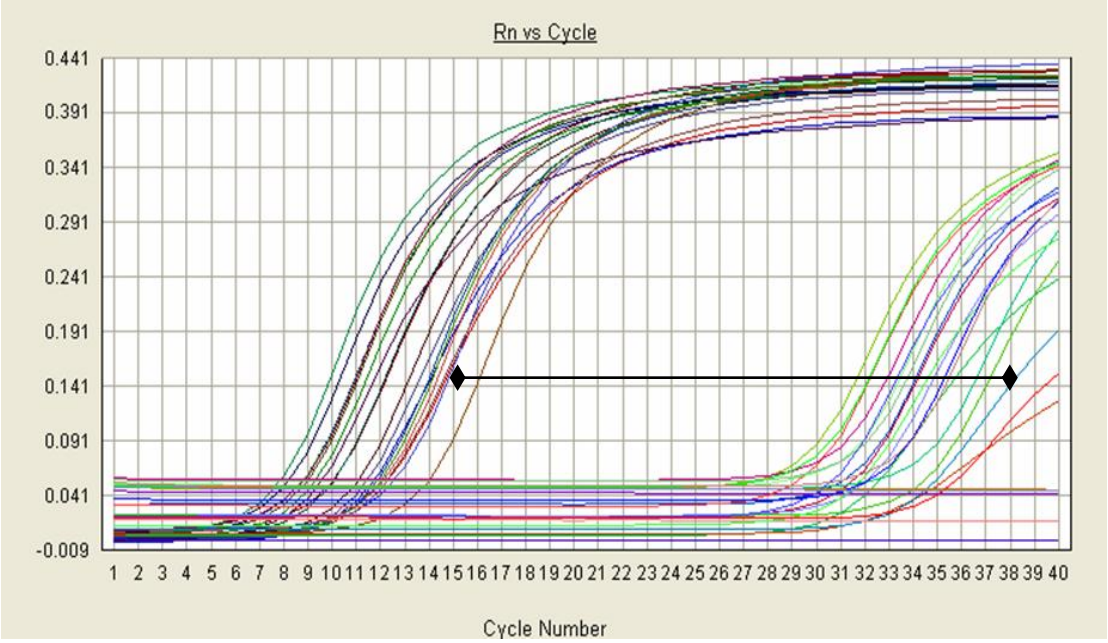


Table 6.2 – Mean cycle at which markers become linear and ΔCt values

	Specific marker			18s Control gene			Delta ct		
	Run 1	Run 2	Mean	Run 1	Run 2	Mean	Run 1	Run 2	Mean
Pax7	34.8	33.8	34.3	12.5	13.2	12.85	21.3	19.5	20.4
NCAM	36.8	37.3	37.05	13.5	12.9	13.2	23.3	24.7	24.0
CD34	32.4	32.6	32.5	12.8	13.8	13.3	18.9	18.3	18.6

Mean values expressed for the individual runs are calculated from the results of the 2 rows of 12 samples (i.e. 24 results) per satellite cell marker.

Expression relative to the control

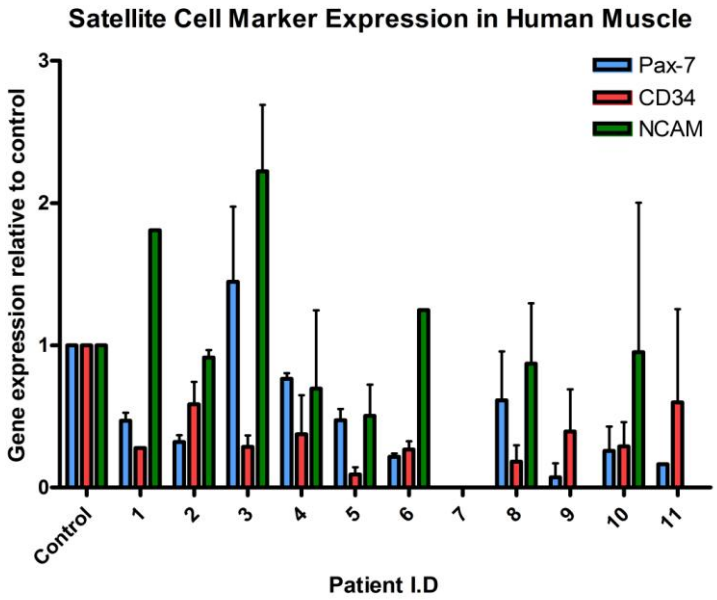
Gene expression of each sample was expressed as a relative amount compared against the control sample, in this case a muscle sample from a young patient (reference to summary Table 6.3 and Figures 6.5 and 6.6). Wide variation in content of the satellite cell markers were found in the samples compared to the control and each other suggesting differing levels of satellite cell content in individual patients confirming the results of the previous experiment. No gene expression was recorded in sample 7 in either run.

Table 6.3 – Gene expression relative to control

Sample	Control tissue	1	2	3	4	5	6	7	8	9	10	11
NCAM	1	1.81	0.88	2.55	0.30	0.35	0.00	0.00	1.17	0.00	1.70	0.00
	1	0.00	0.95	1.89	1.08	0.66	1.25	0.00	0.57	0.00	0.21	0.00
	1	1.81	0.91	2.22	0.69	0.50	1.25	0.00	0.87	0.00	0.95	0.00
PAX-7	1	0.43	0.28	1.07	0.79	0.53	0.23	0.00	0.86	0.00	0.38	0.00
	1	0.51	0.35	1.82	0.73	0.42	0.20	0.00	0.37	0.14	0.13	0.16
	1	0.47	0.32	1.45	0.76	0.47	0.22	0.00	0.61	0.14	0.26	0.16
CD34	1	0.00	0.47	0.34	0.57	0.06	0.31	0.00	0.26	0.60	0.41	1.06
	1	0.28	0.70	0.23	0.18	0.13	0.23	0.00	0.10	0.18	0.17	0.13
	1	0.28	0.58	0.29	0.37	0.09	0.27	0.00	0.18	0.39	0.29	0.60

Data displayed by individual sample for both PCR runs, then mean value (red) below. Note that mean value of the two runs was calculated where two values were obtained. In the absence of a second sample, the value of the single successful run was used.

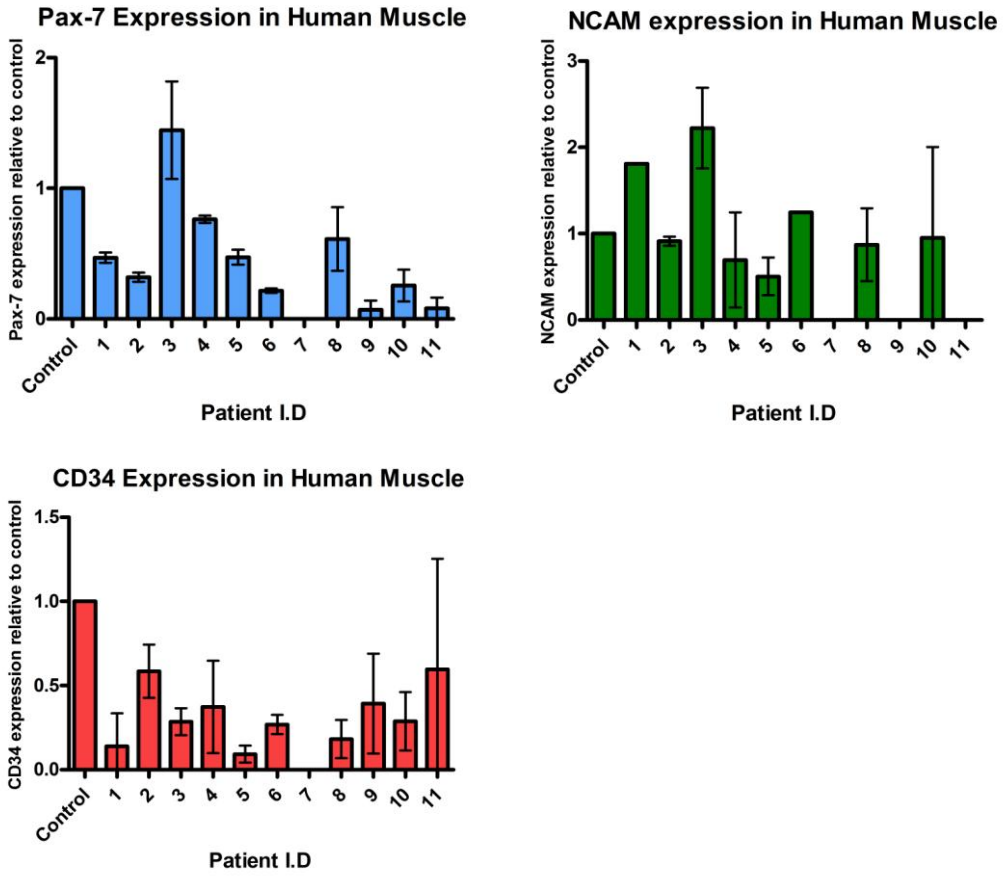
Figure 6.5 - Relative expression of all satellite cell markers. Satellite cell content expressed relative to the control patient. Note the individual variation between the individual samples compared to the control, and also that the 3 markers of satellite cell are present to varying degrees.



All three markers showed considerable individual variation among the samples (Figures 6.5 and 6.6), highlighting different satellite cell content between the

patients. Most of the TKA patients had increased levels of activated cells, and lower levels of quiescent cells than the control.

Figure 6.6 - Individual plots of satellite cell markers and relative gene expression relative to the control sample. No data was obtained on sample 7 suggesting a problem with the processing or storage of that specific tissue that prevented analysis. Pax7 expression (A) demonstrates the expected individual variation in satellite cell content. The NCAM expression (B) generally shows increased levels of activated satellite cells compared to the control sample though was not detected at all in 2 further samples, while the marker of quiescent cells CD34 (C) was consistently expressed less than in the control sample.



Good correlation was observed between the expression of Pax7 and NCAM in the samples ($r = 0.72$, $p = 0.01$). However no correlation existed between the Pax7 and CD34 ($r = -0.03$, $p = 0.93$) or NCAM and CD34 ($r = -0.04$, $p = 0.93$).

Power output

Mean maximal lower limb power output of the group was 16.73 (SE of mean 6.87) W at pre-operative assessment, 35.73 (7.75) W at 6 week assessment and 63.3 (12.1) W at 26 week assessment. This reflected a mean power-body weight ratio of 22.7 (SE of mean 6.8) % at pre-op assessment, 44.6 (8.0) % at 6 week assessment, and 73.6 (9.5) % at 26 week assessment. Wide variation was noted in individual leg extensor power output (Table 6.4).

Table 6.4 – individual patient power output

Sample	Maximal Power Output (W)				Power-Body weight ratio (%)			
	pre-op	6 weeks	26 wks	6-26 wks (change)	pre-op	6 weeks	26 wks	6 -26 wks (change)
1	39	66	135	69	0.3	0.6	1.1	0.5
2	10	37	53	15	0.2	0.7	0.9	0.2
3	72	66	121	55	0.7	0.6	1.1	0.5
4	14	33	58	25	0.2	0.4	0.8	0.4
5	19	23	43	20	0.3	0.3	0.6	0.3
6	30	37	49	12	0.3	0.4	0.6	0.2
7	14	17	38	21	0.3	0.3	0.6	0.3
8	44	83	107	24	0.5	1	1.2	0.2
9	0	18	21	3	0	0.4	0.4	0
10	1	6	57	51	0	0.1	0.6	0.5
11	0	7	14	7	0	0.1	0.2	0.1

Power output for the 11 patients as assessed using the leg extensor power rig pre and post operatively. As in the previous experiment, output is expressed both as maximal power and relative to bodyweight. The important change in power output between 6 and 26 weeks is also recorded.

Change in power output between assessments

Improvement in power output was generally observed between assessment points. Paired samples t-tests were carried out to determine the mean difference between groups between the assessment points. Mean improvement in maximal power output between pre-op and 6 week assessment was 19.0 W. this difference is statistically

significant, 95% CI for mean difference: 3.19, 34.81 p = 0.023. Mean improvement between 6 and 26 weeks was 27.55 W. This difference was statistically significant, 95% CI for mean difference: 13.23, 41.86 p = <0.002.

The improvement in maximal power output reflected an improvement of power-body weight ratio of 21.8% between pre-op and 6 weeks. This difference was statistically significant 95% CI for mean difference: 0.08, 0.35 p = 0.005. A further 29% mean improvement in power to body weight ratio was found between 6 and 26 week assessments, which was also statistically significant 95% CI for mean difference: 0.17, 0.40 p = <0.000.

Again substantial differences between individual scores were noted. Mean improvement of maximal power output between 6 and 26 weeks was 25.6 (6.9) W, which reflected a mean change of 23.6 (5.3) % in power to body weight ratio.

Correlation of satellite cell content and power output

The relative expression of the Pax-7 gene (as the definitive satellite cell marker) was correlated against the change in lower limb power (as per the previous experiment). Good correlation was observed between the change in maximal lower limb power output between 6 and 26 weeks and the expression of pax-7 (r = 0.58, p = 0.06), Figure 6.7, and between the change in power-body weight ratio and the expression of pax-7 (r = 0.79 p = 0.004), Figure 6.8.

Figure 6.7 - Linear relationship found between the patients Pax7 expression and change in their maximal power output post TKA. Good correlation between the variables: r = 0.58, p = 0.06

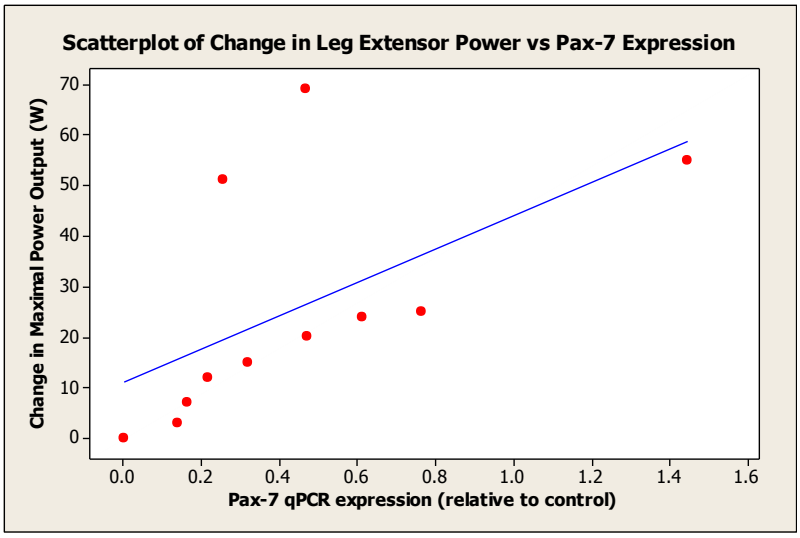
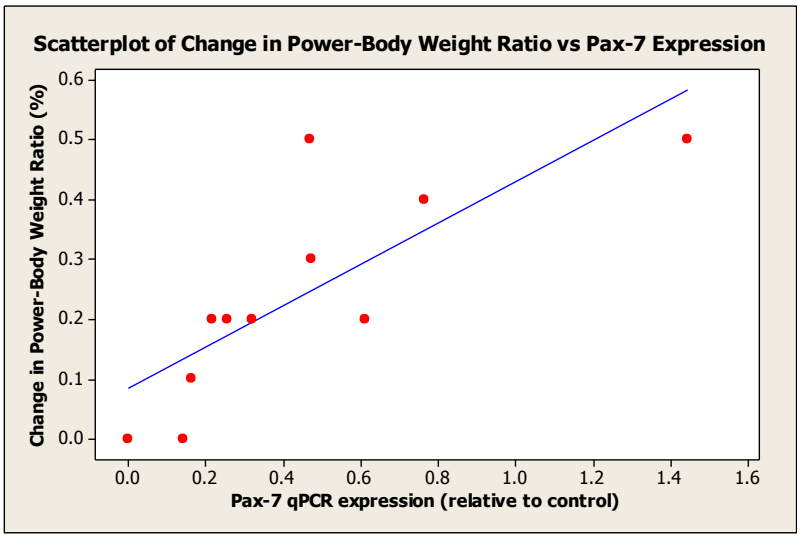


Figure 6.8 - Linear relationship found between the patients Pax7 expression and their change in power-bodyweight ratio post TKA. Strong correlation between variables: r = 0.79, p = 0.004



Correlation with markers of cell activation

Particularly strong correlation was observed between the relative expression of activated satellite cells (detected by expression of NCAM) and the change in the patients maximal power output between 6 and 26 weeks post TKA (r = 0.83, p =

0.002). An equally strong correlation is observed when NCAM was plotted against the patients power- body weight ratio (r = 0.84, p = 0.001) Figures 6.9 and 6.10.

Figure 6.9 - Linear relationship found between the patients Pax7 expression and change in their maximal power output post TKA. Strong correlation between the variables: r = 0.83, p = 0.002

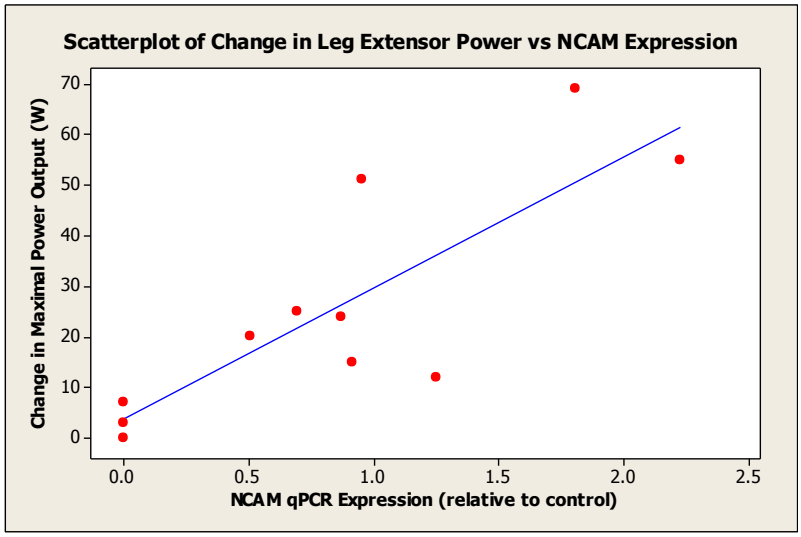
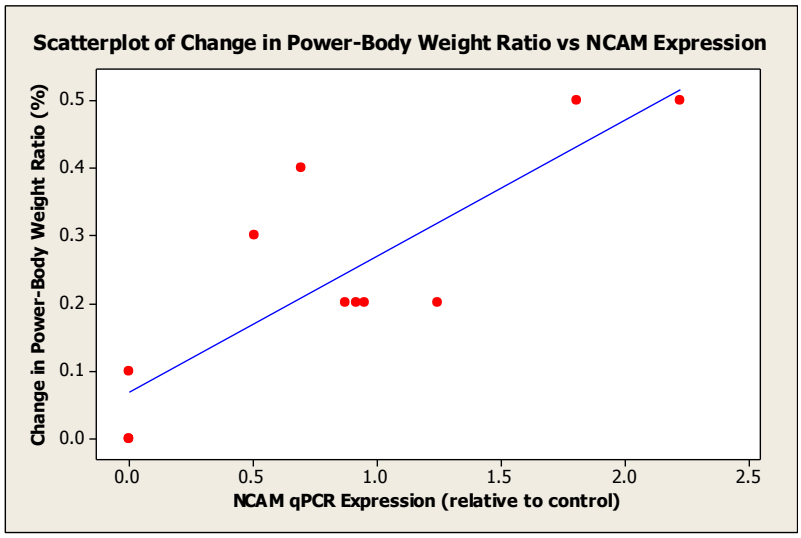


Figure 6.10 - Linear relationship found between the patients Pax7 expression and their change in power-bodyweight ratio post TKA. Strong correlation between variables: r = 0.84, p = 0.001



No correlation was observed between the inactive cells (marked by the expression of CD34) and changes in the patients maximal power output between 6 and 26 weeks post TKA ($r = -0.07$, $p = 0.84$). Equally no correlation was detected when plotted against power-body weight ratio ($r = -0.02$, $p = 0.95$) Figures 6.11 and 6.12.

Figure 6.11 - No correlation was found between the patients CD34 expression and their change in maximal power output, between 6 and 26 weeks post TKA. Correlation between variables: $r = -0.07$, $p = 0.84$

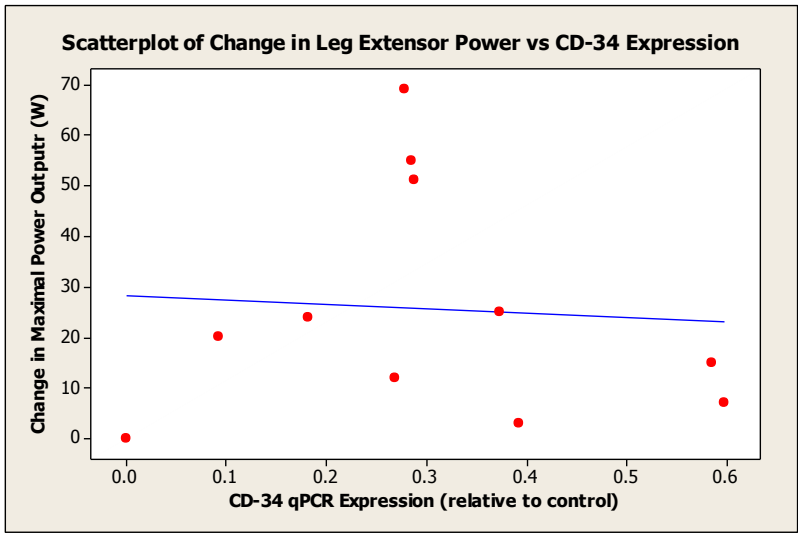
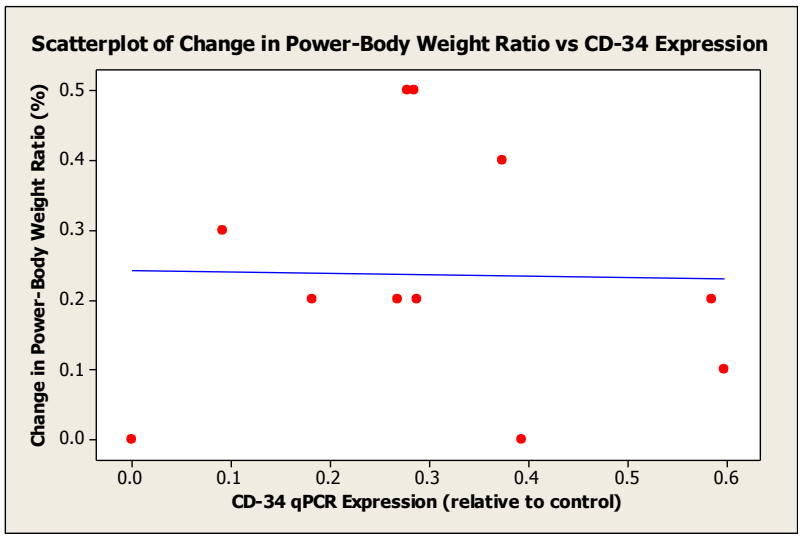


Figure 6.12 - No correlation was found between the patients CD34 expression and their change in power-bodyweight ratio post between 6 and 26 weeks post TKA. Correlation between variables: $r = -0.02$, $p = 0.95$



Regression analysis

Uni-variant regression analysis demonstrated that 26% of the improvement in maximal power output between 6 and 26 weeks could be attributed to the variation in the expression of Pax-7 in the muscle samples (R^2 adjusted 26.0%). Further, 58% of the variation in power-bodyweight ratio could be attributed to the Pax7 expression of the muscle sample (R^2 adjusted 58.2%). Full analysis displayed in Appendix H.

Separately, uni-variant regression analysis demonstrated that 65% of the of the improvement in maximal power output between 6 and 26 weeks could be attributed to the variation in the expression of NCAM in the muscle samples (R^2 adjusted 64.5%). Further, 67% of the variation in power-bodyweight ratio could be attributed to the NCAM expression of the muscle sample (R^2 adjusted 66.7%). Full analysis displayed in Appendix H.

Uni-variant regression analysis demonstrated that none of the variation in improvement in maximal power output between 6 and 26 weeks could be attributed to the variation in the expression of CD34 in the muscle samples (R^2 adjusted = 0.0). Further, none of the variation in power-bodyweight ratio could be attributed to the CD34 expression of the muscle sample (R^2 adjusted 0.0%). Full analysis displayed in Appendix H.

Stepwise regression modelling was performed (alpha to enter or remove p = 0.05) to assess the relative contribution of each factor to the change in power output. The only factor associated with change in post-operative power output (both maximal watts and proportion of body-weight) was the expression of NCAM in the muscle samples (see Appendix H for statistical data output).

6.4 Discussion

This is the first description of quantitative PCR analysis, being employed to assess the content of muscle satellite cells in a clinical population. There are four notable findings of this analysis. Firstly that variation in satellite cell content was found among the patient samples and compared to the control patient (concurrent with the result of the previously presented immunohistochemistry analysis). Secondly, that this variation was correlated with change in power output between 6 and 26 weeks post TKA, in support of the previous experiment. Thirdly, that a stronger correlation with power output was observed with the marker of activated cells than with the generic satellite cell marker, and that this accounted for two thirds of the variance in power output. Fourthly, that that the marker of quiescent cells did not correlate with power output at all.

As with results reported in Chapter 5, the definitive satellite cell marker Pax7 was found to vary among patients, and in this analysis to a control subject (Figure 6.5 and 6.6). This supports the findings of the previous chapter. It is disappointing that the IHC analysis attempted in this chapter was unable to determine a satellite cell staining profile that would enable direct comparison to the previous work. Irrespective, the qPCR analysis confirmed the previous conclusions of varying satellite cell content in the quadriceps muscle of TKA patients, and that this correlated to physical recovery (Figures 6.7- 6.10). The primary focus of this experiment was to quantify the amount of satellite cell markers in the biopsy using a sensitive method of analysis; this has been achieved.

The Pax7 marker correlated well with maximal lower limb power output ($r = 0.58$) and very well with power-bodyweight ratio ($r = 0.79$), broadly in line with the previous results. Uni-variant regression analysis demonstrates that the cell content accounts for around 26% of the variance in maximal power output (which was remarkably similar to the findings presented in Chapter 5), but in excess of 50% of

the variance of change in lower limb power output when scaled for bodyweight. The NCAM marker correlated very well with both maximal lower limb power output ($r = 0.83$) and with the power-bodyweight ratio ($r = 0.84$). Uni-variant regression of this demonstrates that around 66% of the variance in both maximal lower limb power output and power-bodyweight ratio could be attributed to the underlying satellite cell content (Appendix H).

Stepwise linear regression modelling demonstrated that the variation in the expression of NCAM in the muscle samples alone was the strongest predictor of variation in power output both in terms of maximal watts, and proportional to bodyweight. These proportions were substantial and potentially explained most of the differing physical outcome among knee replacement populations.

While this figure is far larger, than was suggested in the preceding chapter, the technique used to determine the results was more sensitive and reliable as all of the tissue was used for analysis, as opposed to thin sections. Specific genetic sequences were used to detect the amount of the markers in question, and the amount found per sample is likely to be correct. The findings would be strengthened by an increase in the size of the cohort of TKA patients from which physical data was collected, due to the large standard deviations noted for this assessment in previous chapters. The power output changes (Table 6.4) were however largely comparable to those of the previous cohort (Table 5.2) and this lent further support to the conclusions of this chapter.

The finding that the quiescent cells do not correlate with the power output is not surprising. However the expression of CD34 detected was greater than that of Pax7, suggesting that there was more of the CD34 marker in the samples. This was not expected, but could be due to CD34 expression not being specific for satellite cells.

CD34 is a cell surface glycoprotein and functions as a cell-cell adhesion factor that has been used previously as a marker of quiescent satellite cells (Beauchamp et al, 2000). Neilson and McNagny (2008) however have demonstrated expression of this factor on early haematopoietic and vascular associated tissue. As such it is probable that the content reported incorporated some of this other tissue, and was not unique to quiescent satellite cells. As there was no correlation between the content of the CD34 marker with either Pax7 or with NCAM, common levels of over expression cannot be assumed, and the correlation between quiescent cells and power output should be interpreted cautiously. Good correlation was detected between Pax7 and NCAM content in the samples ($r = 0.72$) and this supports the use of these markers.

In conclusion, these results suggest strong correlation between satellite cell number at the time of surgery and subsequent change in power output post operatively. They largely corroborate the results presented with a differing analysis technique in Chapter 5. It was further found that the expression of the marker of the satellite cells that were activated, as opposed to being quiescent, was most strongly correlated with the change in post-operative lower limb power output, and potentially explains around two thirds of the variation in change in post-operative power output.

The results presented support the hypothesis that the size of the individual's satellite cell pool and the number of activated cells within are directly relevant to the subsequent recovery of muscle power and thus physical function following TKA.

7 Conclusions and Future Directions

7.1 Conclusions

The aim of this thesis was to determine whether mechanically advantageous prosthetic design and / or the regenerative capacity of the patient's muscle influenced the patient's physical function following total knee arthroplasty. This was specifically investigated by assessing power output of the extensor mechanism.

A randomised controlled trial was conducted to compare a new knee prosthesis designed with a single radius of curvature femoral component, which suggested beneficial muscle function through its design, with a traditional 'multi-radius' knee replacement model. Patients with the new implant design reported superior outcome at 1 year (as demonstrated by the Oxford Knee Score) and specifically demonstrated enhanced lower limb power output. The time course of improvement in power output over the first year post-operatively was also enhanced in patients with the new implant compared to the control group.

Muscle satellite cells isolated from biopsies of the quadriceps muscle of patients at the time of surgery accounted for a third of the variance of the change in power output post operatively. Activated satellite cells were found to account for around two thirds of the change in post-operative power output.

A further research question was to assess the relationship between patient reported outcome and direct functional evaluation. This was done to enable comment on the ability of the patient report tools to identify changes in the physical performance of the individual, the capacity of which is currently debated. This allows further interpretation of the changes in extensor mechanism power detected within the context of overall patient post-operative outcome.

Analysis of direct testing of outcome and patient reported outcome demonstrated a changing relationship between function and patient report of that function over time, with a closer relationship between the two types of assessment as the influence of pain diminished post-operatively. This confirmed that patient reported assessments are not equivalent to direct physical evaluation, and suggests that specific functional testing is required to assess the magnitude of the influence of mechanical and physiological factors investigated in this thesis. Correlation of this specific assessment can then be drawn to overall functional outcome as measured by patient report methods. The clinical assessment model derived suggests a means by which to explain this broader relationship through levels of the implant, the limb and the person.

Mechanical influence

A new design of total knee arthroplasty femoral component that is hypothesised to mechanically advantage the extensor mechanism of the knee was introduced in Chapter 2. Clinical demonstration of this theoretical advantage had not been previously been confirmed.

A double blind randomised controlled trial of 212 TKA patients to compare the new implant design with a traditional model was presented in Chapter 3, where patient outcome was assessed at four time points over a one year period. Patient outcome at one year was superior in measures of knee flexion, lower limb power output and by patient report questionnaire (Oxford Knee Score), Two-way ANOVA, $p = <0.001$ in all cases. Specifically, the extensor mechanism power was significantly increased between all four assessment points in the new implant group, the control group demonstrating change between the second and third assessment only ($p= <0.001$), as was predicted by the implant design.

Many factors relating to both the patient and the surgeon are thought to influence post-operative patient function. In this investigation, randomisation of the study participants accounted for factors relating to the patient, while the use of a limited number of orthopaedic surgeons at the same institution and adherence to standardised local protocol for post-operative management and rehabilitation limits any affect due to the surgeon. Controlling for these known variables allowed investigation of any specific effects based on differing prosthesis implanted.

The influence of the improved lower limb power output found with the new single radius design was considered in Chapter 4. Patient outcome in the wider literature is typically assessed with patient reported questionnaire methods, the merits of which were considered in Chapter 2. Patient reports of their levels of function were found to differ to direct assessment of functional ability, and the influence of pain on patient report of function was highlighted. The relationship between patient report of function and direct assessment of function was found to change over time; the worst association was found prior to surgery, while post-operatively an improvement in the relationship was found at each sequential assessment at six weeks, six months and twelve months post TKA. The clinical assessment framework presented suggested that the influence of pain was most apparent at the level of the person, and that separate assessment at the level of the limb was required to ascertain specific information relating to physical performance.

Assessing global outcome by patient report questionnaires alone allowed other factors relating to the patients functioning in society to affect the overall report of function. In this context the substantial improvements in extensor mechanism power found in the single radius implant group were perhaps not fully reflected in the corresponding Oxford Knee Scores. Despite this potential limitation, statistically significant differences were found in OKS between the implant groups at a magnitude that indicated clinically relevant differences.

Physiological influence

The second specific research question concerned the role of the intrinsic number and activation state of muscle satellite cells present in the patient's quadriceps muscle on the patient's post-operative recovery of muscle power. The known association between muscle power and patient outcomes following TKA was presented in Chapter 2. The mechanisms by with satellite cells differentiate to provide new myoblasts to facilitate hypertrophy or hyperplasia of muscle tissue was also reviewed. The physical recovery that follows TKA was hypothesised to be dependant on the intrinsic number of satellite cells, though the lack of clinical investigation of human muscle satellite cells was also highlighted.

A pilot study was conducted to determine if the number of satellite cells in the quadriceps muscle influenced post-operative recovery. Muscle satellite cells were isolated from biopsies of the quadriceps muscle of 18 patients at the time of surgery and counted by an immunofluorescent staining technique was presented in Chapter 5. The number of satellite cells detected accounted for a third of the change in power output post-operatively (R^2 = 36.6%), and was demonstrated to explain 10 times the variance of post operative power output than the patient's pre-operative power output values. The wide variation found in individual power output and strong relationship to the muscle satellite cell number in this patient cohort (r = 0.64) suggests the importance of these cells in post-operative recovery. A larger sample size is required to ascertain the implications of this pilot work, as is data concerning the cell numbers in a healthy control population.

The wide standard error of the mean of the power data in this cohort was due to the large variation in power output among the individual patients. Though partly compensated for by the use of change in power output and by scaling the power relative to the patient's body weight, larger numbers of patients would be required to reduce the error of the mean of the power output data. The wide variation seen in the standard error of the mean of the power output of the 200 patient's assessed in

Chapter 3 suggested that a substantial volume of patients may need to be investigated before definitive conclusions could be drawn. Despite this, normal distributions were found in both maximal power scores and of the change in power scores, which allowed parametric testing of between time point differences and for linear correlation between variables.

General function was not specifically assessed in this pilot work, though previous studies of large numbers of patients have shown a relationship between muscle power and post-operative function (Faulkner et al, 2010; Lingard et al, 2004). Additionally the modest - good correlation and linear association between lower limb power and both functional assessment and patient report of their function (through the Oxford Knee Score) found in Chapter 4, suggested that an improvement in lower limb power output would be indicative of improvement in other functional parameters. Again, the difference found in functional assessment and patient report of function in Chapter 4 suggested that no one outcome measure is suitable to ascertain overall patient function.

A preservation of the muscle satellite cell pool has been previously shown in elderly subjects who undertake regular exercise. It is possible that typical osteoarthritic patients exercise less than the healthy population due to the pain associated with the disease and thus may have less well preserved satellite cell pools. This may then limit muscle regeneration post-operatively, and subsequently physical function. This is an attractive explanation for the continued muscle power deficits found in long term post-operative follow-up studies compared to the patient's contralateral limb and to healthy controls.

Confirmation of the relationship found between satellite cell number and change in post-operative muscle power output was presented in Chapter 6 in a separate cohort of 11 patients. This analysis was conducted using a quantitative PCR technique that was more sensitive than immunofluorescent staining as it quantified the expressed DNA of the markers of the satellite cells. This was the first example of this technique

being employed to detect satellite cells in human samples. It was found that the activated satellite cells accounted for twice the variation of change in post-operative power compared to the generic marker of satellite cells which would include those in a quiescent state. The activated cells explained around two thirds of the change in post-operative power output (R^2 = 66.7%). The assessment of larger numbers of patients is required to confirm these exciting results.

7.2 Future directions

Further development of mechanically advantageous implant designs may be beneficial; however the mechanical advantage investigated in this thesis was elicited by placing the flexion axis more posteriorly in the knee. Attempting to locate this axis further posteriorly would in principle be possible, though the current implant design was based on a new kinematical theory of knee motion. Moving the axis beyond this location (as would be required to enhance the moment arm further) would potentially create another situation of a multi-instantaneous centre of rotation that has been criticised, and led to the development of the single axis design.

An alternative assessment to the suggestion that the single radius Triathlon implant design mechanically advantages the extensor mechanism is that previous multi-instant centre of rotation designs may have mechanically disadvantaged the extensor mechanism function. This is an attractive idea that may offer some explanation as to why some patients' who were previously able to complete activities of daily living can struggle to recover functionally following knee arthroplasty, despite a technically well located and well fixed implant. Further work is needed to confirm the findings of the randomised trial and also to assess the wider relevance of enhanced quadriceps function within the context of overall patient function following total knee arthroplasty.

An additional priority for future research in this area would be the comparison of a 4th generation multi-radius implant design to the single radius Triathlon implant. This would determine if any of the additional design modifications beyond the axis of rotation are relevant to the outcome of the trial presented here.

The exciting results of the satellite cell studies suggest great potential for further research, though the first priority would be to confirm the results presented here in a larger sample of patients.

The results presented in this thesis suggest that the number of satellite cells within the individual's cell pool and the activation state of these cells are highly relevant to the recovery of physical function post TKA. The mechanisms involved in the regulation of skeletal muscle growth and regeneration are of great interest, as through the therapeutic manipulation of these mechanisms it may be possible to enhance a patients muscle recovery and the quality of life. Spangelberg and Booth (2001) comment that in the future it may be possible to regulate the proliferation of the satellite cells either via gene delivery to the skeletal muscle, or via isolation of the satellite cells, genetic manipulation and delivery back to the muscle via the circulation. Substantial further advances in biological technology are however required to achieve this goal.

Perhaps a more immediate solution will emerge from the manipulation of the host environment to induce quiescent satellite cells into an activated state. Hall et al (2010) recently achieved this in a murine experiment and demonstrated substantial physical muscle hypertrophy compared to the contralateral limb when injury was induced. Corroboration of these results in human populations is first required and it is likely that a large volume of experimental work will be required to optimise these cell protocols due to the known differences between mouse and human satellite cell markers.

Further work however is needed before any clinical benefits could be expected using therapies such as modifying the satellite cell environment. Identification of those patients at most risk of poor physical recovery would be essential to make this technique a realistic therapy. It would be preferable not to have to biopsy the muscle in order to determine the satellite cell content. Clinical algorithms of 'at risk' patients could be developed and these patients at a higher risk of having poor intrinsic muscle regenerative capacity could then have their satellite cell content confirmed with a biopsy.

Despite these future hurdles, the demonstration that post-operative muscle function can be influenced by manipulating the quadriceps / extensor mechanism function, both mechanically by implant design, and physiologically by the intrinsic 'regenerative potential' of the muscle tissue (by virtue of the volume and activation state of the satellite cell) suggests the potential to enhance patient outcomes following total knee arthroplasty.

8 Appendices

Appendix A: Submitted paper

Hamilton DF, Henderson GR, Gaston P, MacDonald D, Howie C, Simpson AHRW. Outcomes after total hip arthroplasty are superior to outcomes after total knee arthroplasty: a prospective cohort study

Revision submission at editor's request, following feedback from reviewers

Under review at time of thesis submission.

Outcomes after total hip arthroplasty are superior to outcomes after total knee arthroplasty: a prospective cohort study

Journal:	*BMJ Open*
Manuscript ID:	BMJ Open-2010-000015.R1
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Hamilton, David; Universtiy of Edinburgh, Orthopaedics and Trauma Henderson, Robin; NHS Lothian, Department of Medicine of the Elderly Gaston, Paul; University of Edinburgh, Orthopaedics and Trauma MacDonald, Deborah; University of Edinburgh, Orthopaedics and Trauma Howie, Colin; University of Edinburgh, Orthopaedics and Trauma Simpson, Hamish; University of Edinburgh, Orthopaedics and Trauma
Subject Heading:	Evidence-based practice
Keywords:	hip arthroplasty, knee arthroplasty, Orthopaedics

ABSTRACT

Objective: To investigate the comparative patient reported outcome of total hip and knee arthroplasty in the first year following surgery

Methods: This prospective study included all elective primary total joint arthroplasty procedures (1410 hip and 1244 knee) performed at the Royal Infirmary of Edinburgh over a 2 year period, between January 2006 and November 2008. Patient reported outcome questionnaires for general health (SF12) and joint specific function (Oxford Score) were completed pre-operatively and at 6 and 12 months post-operatively. Overall satisfaction was assessed at 12 months post-op.

Results: Both groups demonstrated substantial improvement in the year following surgery, though the THA group demonstrated greater improvement than the TKA. On the Oxford Scores, THA outcome improved on average by 4.9 points more (95% confidence interval [4.2, 5.7]) than TKA at 6 months and by 4.4 [3.7, 5.1] points at 1 year. The SF12 physical scores were on average 2.6 [3.4, 1.8] points better than TKA at 6 months and 2.7 [3.5, 1.9] points better at one year. SF12 mental health scores varied little throughout the post-operative year. Analysis of covariance reveals that the type of arthroplasty and pre-operative Oxford Score are predictors of outcome.

Conclusions: Both procedures confer substantial improvement in patient outcome one year following surgery; however greater joint specific, general health and satisfaction scores are reported following THA and a faster rate of improvement is demonstrated compared to TKA.

Article focus:
Common perception is of equivalent outcome following hip and knee arthroplasty

Mean changes in patient reported scores are quantified for both hip and knee arthroplasty.

Models that may be used to predict mean scores for patients with given pre-op Oxford Score were developed.

Key messages:
THA patients report superior outcomes and faster improvement than TKA patients.

This difference in outcome is predominantly physical in nature, and mostly occurs in the first 6 months.

The extent of the difference in outcome may be informatively displayed via our models.

Strengths and limitations:
This is the largest and most comprehensive study to assess this issue.

The models must be considered with some caution due to the discrete and bounded nature of the Oxford Score. In addition they explain only a fifth of the variation in outcome.

INTRODUCTION

Total hip and knee arthroplasty (THA and TKA) are very common procedures (each in excess of 70,000 per year in UK) that are highly successful in treating the morbidity of patients with end-stage arthritis[1-3].

Outcomes of both procedures are commonly perceived by patients and clinicians to be equivalent. This has been reinforced in the general medical literature, Gidwani and Fairbank[4] writing specifically about TKA noted comparable outcome with THA as a BMJ article summary point.

A variety of cohort studies with small numbers of patients and using differing generic measures of health outcome have directly compared the outcome of the two procedures and reported conflicting results[5-10]. Wylde et al[11] conducted a 'mid-term review' comparing THA and TKA at a single centre using only joint specific outcome measures. Oxford Hip and Knee scores were directly compared post-operatively to suggest improved comparative function following hip replacement. Pre-operative data was not available, thus comparative change in scores could not be addressed. The originators of the Oxford Scores however consider this analysis inappropriate as the separate hip and knee questionnaires do not ask equivalent questions and suggest the results may be misleading[12].

The aim of this study was prospectively to compare the patient outcome of total knee and hip arthroplasty in a large cohort, in the first year following surgery, using relevant standardised instruments to form comprehensive assessment of general health, joint specific outcome and overall patient satisfaction. This information will help inform patients and healthcare providers, who are involved with both referrals into orthopaedic services and with subsequent post-operative management, as to the expected outcomes of the surgery.

There are two valid approaches to assessing outcome from pre and post operative data, the use of change score (difference in post-operative score from pre-operative values) or to use the pre-operative data as a covariate in analysing the final outcome. These approaches represent two separate research questions, the first asking whether there is a difference in average change of the two populations, the second asking whether a member of group 1 is expected to change more than a member of group 2 if they have the same initial value[13]. We incorporate both these questions into our analysis.

MATERIALS AND METHODS

We prospectively followed all elective primary total hip and primary total knee arthroplasties performed at the Royal Infirmary of Edinburgh in the 2 years between January 2006 and November 2008 inclusive. This reflected some 1410 total hip and 1244 total knee procedures.

The outcome assessments used were the Oxford Hip or Oxford Knee Score, the Short Form 12 (SF-12) and a separate validated satisfaction question[14]. Patients were asked to complete these self-administered questionnaires at the time of pre-operative assessment and then by postal survey at 6 and 12 months post operation.

1410 THA and 1244 TKA datasets were available for analysis pre-operatively, 1389 THA and 1223 TKA at 6 months and 1381 THA and 1227 TKA at 12 month follow-up. This represents a loss to follow-up of 2% THA and 1.4% TKA patients in the year post surgery.

Satisfaction data was recorded at the 12 month follow-up, 1348 THA and 1185 TKA patient data were available. All data was included in the analysis to limit any effect of loss to follow up[15].

The Oxford Hip or Knee Score results in a single outcome score between 12 and 60 (a high score indicates increased levels of disability because of pain and poor function, with a reduction in the Oxford Score indicating improvement). Evaluation of comparative change in the respective hip and knee scores was performed to compare between procedures. The SF-12 results in two scores, the physical and mental components (MCS and PCS). Its scoring is based on norm-based methods using population mean scores. Both PCS and MCS have a population mean score of 50, with standard deviation of 10. Higher scores denote better function with these instruments[16]. The satisfaction question consists of a 4 point Likert scale, with answers ranging from dissatisfied to very satisfied.

Statistical Analysis

The first analysis performed addressed the change-score approach i.e. does the average change in patient report scores differ between the two populations? Fitzmaurice (2001)[17] refers to this as an *unconditional* research question that compares the average (or unconditional mean) change-score in one population with the average change-score in a second population. The analysis involved use of ANOVA via the GLM facility in Minitab.

The second analysis (analysis of covariance or adjusted change-score analysis) essentially addressed what Fitzmaurice[17] refers to as the *conditional* research question: Is there any difference between the expected change-score of a THA patient and TKA patient if they have the same baseline score? In this analysis the opportunity was also taken to investigate the potential influence of the covariates age and gender, as these are thought to influence individual outcome of joint arthroplasty. This analysis was performed using multiple regression methods. All data analysis and display was carried out with the Minitab (Release 15) software.

RESULTS

Summary statistics

Differences in age profiles of the groups were assessed with independent sample t-tests. THA patients were 2 years (95% confidence interval [-2.86, -1.33] p = <0.001) younger at 68.1 years at time of operation compared to TKA patients at 70.2 years. The male: female ratio in both groups was similar, and no significant difference was found between the mean ages of male and female patients within hip and knee groups (Table 1).

Table 1 – Age at time of surgery

	Hip (n = 1410)		Knee (n = 1244)	
	Age (Mean+/-SD)	% patients	Age (Mean +/-SD)	% patients
Male	67.8 (11.0)	42.8%	70.0 (8.8)	43.2%
Female	68.4 (11.0)	57.2%	70.4 (9.3)	56.8%

Table 2 displays means and standard deviations of the patient reported outcome measures. The means are also displayed in figures 1 and 2 with 95% confidence intervals for the corresponding population means.

Table 2 – Outcome Scores for Hip and Knee groups, Mean (SD)

	Pre-op		6 months		12 months	
	Hip	Knee	Hip	Knee	Hip	Knee
PCS	29.5 (8.4)	30.2 (10.4)	42.4 (12.6)	39.5 (11.6)	43.2 (13.0)	40.3 (12.4)
MCS	49.4 (12.2)	51.4 (12.1)	53.3 (11.3)	51.8 (11.6)	52.5 (11.5)	51.7 (11.6)
OXS	41.6 (8.3)	41.3 (7.6)	22.4 (9.2)	27.4 (10.1)	21.4 (9.5)	25.8 (10.3)

Figure 1 - Mean Oxford Scores

Figure 2 - Mean SF-12 Scores

Analysis 1

An ANOVA was performed via GLM in Minitab with factors operation (levels THA and TKA), patient (nested within operation) and occasion (with levels pre-op, 6-month and Year). An operation-occasion interaction term was included in the model and 95% Bonferroni confidence intervals (displayed in Table 3 and 4) obtained for all possible operation – occasion pairings.

Table 3 – Within operation comparisons

			Change in mean score	95% CI
OXS	Pre-op – 12 months	Hip	-20.2	(-20.9, -19.5)
		Knee	-15.4	(-16.2, -14.7)
	Pre-op – 6 months	Hip	-19.2	(-19.9, -10.5)
		Knee	-13.9	(-14.6, -13.1)
	6 months – 12 months	Hip	-1.0	(-1.7, -0.3)
		Knee	-1.5	(-2.3, -0.8)
PCS	Pre-op – 12 months	Hip	13.2	(12.4, 14.0)
		Knee	10.2	(9.4, 11.0)
	Pre-op – 6 months	Hip	12.3	(11.6, 13.1)
		Knee	9.4	(8.6, 10.2)
	6 months – 12 months	Hip	0.9	(0.1, 1.6)
		Knee	0.8	(-0.02, 1.6)ns
MCS	Pre-op – 12 months	Hip	3.1	(2.1, 4.1)
		Knee	0.3	(-0.8, 1.3)ns
	Pre-op – 6 months	Hip	3.9	(2.9, 4.9)
		Knee	0.6	(-0.6, 1.5)ns
	6 months – 12 months	Hip	-0.8	(-1.8, 0.2)ns
		Knee	-0.2	(-1.2, 1.5)ns

Table 4 - Between operation comparisons (Estimate of μ_{Knee}-μ_{Hip})

	OXS		PCS		MCS	
Occasion	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Pre-op	-0.4	(-1.1, 0.3) ns	0.3	(-0.5,1.1) ns	2.0	(1.0, 3.0)
6-month	4.9	(4.2, 5.7)	-2.6	(-3.4, -1.8)	-1.5	(-2.5, -0.5)
Year	4.4	(3.7, 5.1)	-2.7	(-3.5, -1.9)	-0.8	(-1.8, 0.2) ns

This analysis provides evidence, via these confidence intervals, that mean Oxford score continued to improve significantly over the year for both procedures. Mean PCS improved significantly for both procedures in the initial 6 months. Small further improvements were seen between 6 months and 12 months, though were not significant in the TKA group. Pre-operatively the differences in mean MCS were statistically significant for the two groups, with TKA patients having the higher mean. This situation was reversed at 6 months. There was no significant difference after a year.

As an example of a between operation comparison, one year post arthroplasty the mean reduction in Oxford Score for hip patients is 4.4 points greater than that for knee patients with 95% confidence limits 3.7 and 5.1 points.

Analysis 2

In order to carry out this analysis multiple regression was employed with Oxford score, at either 6 months or a year, as the response variable, with type of arthroplasty and time as factors, and pre-operative Oxford Score, age and gender as covariates. Variable selection procedures yielded a model involving operation and pre-op score with an adjusted R-squared value of the order of 20%. It was found that inclusion of age and gender did not improve the explanatory power of the models. Logarithmic transformation of the responses yielded residual plots that were more satisfactory than those obtained using the untransformed response but did not have any impact on the R-squared values. Because of the discrete and bounded nature of the Oxford Scores the models must be treated with some caution. George Box's dictum "All models are wrong, but some are useful!" should be borne in mind[18].

The model displayed in figure 3 is the ANCOVA model with the Oxford Score one year post-op as the response. For example it predicts that THA patients with pre-operative Oxford Score of 40 would have mean Oxford Score of the order of 19 points whereas TKA patients pre-operative Oxford Score of 40 would have mean Oxford Score of the order of 23 points. Readers are cautioned that there is wide variability in individual patient outcomes as indicated by the underlying scatter plot.

Figure 3 – Scatter plot with ANCOVA model for Oxford Score 1 year post-op

Figure 4 – ANCOVA models for Oxford Score at 6 months and 1 year post-op

Figure 4 indicates that, taking the same example of a patient presenting with a pre-operative Oxford Score of 40 points, on average, THA patients improve by around 1 further point between 6 months and 1 year post-op, whereas TKA patients improve by around 2 points.

Satisfaction with outcome

Very high levels of patient satisfaction were recorded for both procedures. However the proportions of patients recording overall satisfaction with the outcome of the arthroplasty at one year was significantly greater for the THA group (91.1%) than the TKA group (81.4%) (p-value <0.001). Thus 18.6% of TKA patients, around twice the proportion of THA patients (8.9%), did not consider themselves satisfied with the result of the procedure one year following surgery.

DISCUSSION

We are aware of no other study that has prospectively assessed a large cohort of hip and knee arthroplasty patients at a single centre in order to compare the patient outcomes of the two procedures directly.

Hip arthroplasty was found to outperform knee arthroplasty as measured by relative change in the Oxford Hip / Knee Scores, SF-12 score and level of patient satisfaction.

When assessing outcome it has been suggested that a combination of a joint specific and general health assessment tool offers the best combined analysis. The Oxford Hip and Knee Scores and the Medical Outcomes Study Short Form-12 questionnaires are highly validated and reliable tools that are accepted by patients and surgeons to gauge pain and functional outcome[19-21].

We found the difference in outcome between THA and TKA groups to be predominantly physical in nature. At 12 months post-op mental health scores (MCS) were equivalent between the THA and TKA groups, whereas physical scores (Oxford Score and PCS) were significantly worse in the TKA group. The detected significant differences in MCS between our operative groups at pre-op and at 6 months post-op were small in magnitude, within one standard deviation of the population mean, and unlikely to be clinically relevant.

Physical outcome scores for both procedures follow the same trend of substantial improvement in the first 6 months following surgery, and then subtle further improvement in the second 6 months (figures 1 and 2 and table 3). It is in the initial 6 month period where the majority of difference in comparative improvement between the hip and knee procedures occurs. The small spontaneous improvement between 6 and 12 months suggests that patients functioning poorly at 6 months should perhaps be considered for referral to targeted therapy at this point. Both groups demonstrated similar improvement in average scores in the second 6 months (figure 4). Of further interest is that while the mental health scores remain around normative values throughout, the physical scores for each group are initially very low then improve dramatically following both procedures, but do not reach normative values on the SF12 score.

Previous authors have attempted to assess comparative post-operative function by way of generic health measures. Ritter et al[6] assessed 85 THA and 93 TKA patients in terms of quality of life and general health by the SF-36 questionnaire and reported no difference in

results. Benroth et al[9] reported on 63 hips and 110 knees and were also unable to detect a difference using SF-36 as an outcome measure. Norman-Taylor[7] reported a small cohort of 41 THA and 31 TKA and suggested similar outcomes were achieved, however in order to compare the different outcome scores, they carefully converted the Harris Hip scores and a modified British Orthopaedic Association Knee functional assessment chart into Rossiter distress and disability scores and then compared them on the Rossiter index matrix. From this they derived quality of life scores but found no significant difference between the operative groups.

Conversely Bachmeier et al[8] reported significantly enhanced WOMAC and SF-36 scores for hip arthroplasty patients (n = 86) at 6 and 12 months post-operatively compared to knee arthroplasty patients (n = 108). In a review Ethgen et al [1] commented on the conflict in the literature concerning the results of THA and TKA. They considered that for quality of life outcome measures patients did better following THA, however meta-analysis was not performed. Bourne et al[10] assessed a large Canadian cohort using the WOMAC score and willingness to undergo the procedure again as outcome measures. They suggested superior outcome with THA, but their study suffered from a 30% loss to follow-up. O'Brein et al[5] suggested improved function in hip patients using a comparison of change in the joint specific Oxford Hip/Knee Scores as the sole outcome measure. Wylde et al[11] also reported superior outcome of THA compared to TKA using the Oxford Scores. Their analysis was limited however by not presenting any pre-operative data. Further Dawson et al[12] have criticised the methodology of Wylde's paper, highlighting the different population characteristics of hip and knee patients and that 3 of 12 questions are different in the respective Oxford Hip and Knee Score questionnaires. Dawson et al[12] commented that direct comparison of the respective mean Oxford Scores is not valid and thus the results suggested by Wylde et al[11] potentially misleading.

We concur that directly comparing the means of the Oxford Hip and Oxford Knee Scores is controversial but consider that comparing the relative change in the scores to be of value, particularly as our operative groups had very similar distributions of baseline Oxford Score. The comparative change of approximately 5 points on the Oxford scale is relatively large and this magnitude of change on either hip or knee score would be considered clinically significant.

Of particular interest was that the overall patient satisfaction with the procedure also differed between hip and knee arthroplasty, with more patients in the THA group being satisfied (91.1%) compared to the TKA group (81.4%). This supports the finding of superior results after hip arthroplasty suggested by SF-12 scores and change in Oxford Scores.

We have attempted to answer two separate research question in our analysis. The first relates to the broad question of overall group change and the second sought to take into account any important covariates that may influence outcome. Interestingly, both analyses provide the same answer of approximately 5 point greater improvement in the Oxford Score in the THA group compared to the TKA group. This similarity in output results from the very similar baseline values of the two groups and the lack of effect of age and gender on the covariant analysis.

Our model identifies that the procedure factor (operation) and baseline score covariate are predictors of outcome. However, there is a large proportion of unexplained variation. Current work within he orthopaedic community seeks to create more effective models for the prediction of post-operative outcome. Statisticians are also addressing issues associated with

discrete and bounded responses such as the Oxford Scores[22]. Our model can be seen as a useful starting point with which to inform both patients and clinicians as to the likely average outcome of each procedure.

In conclusion, our results from a large cohort demonstrate that there is a high level of satisfaction and considerable improvement in patient reported outcome measures after both total hip and total knee arthroplasty. Patients are however more likely to report a greater improvement following total hip arthroplasty compared to total knee arthroplasty. Most improvement in patient scores occurs in the initial 6 months period post-op, however small further improvements in physical function can be seen between 6 and 12 months post-op.

Funding
Stryker UK provide funding to the Department of Orthopaedics and Trauma, University of Edinburgh to support the patient database. They had no role in the design, collection, analysis or interpretation of the data, or in the writing of the article or decision to submit for publication

Competing Interests
DH is supported by a PhD studentship from the Medical Research Council doctoral training scheme and Stryker UK, none of the other authors declare competing interests.

Data Sharing
Dataset available form the corresponding author though subject to approval of the data manager due to NHS restrictions in place to protect patient confidentiality.

Contributors
All authors were involved in the conception and design of the study protocol. DM monitored collection and collated the data. DFH and GRH performed the analysis and drafted the manuscript. All authors contributed to and approved the final manuscript. AHWR is the guarantor.

Ethical approval
Approval was granted by the Lothian Research Ethics Committee, all patients consented to having their data included on the database and for this to be used and published for research proposes

REFERENCES

[1]Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY**.** Health-related quality of life in total hip and total knee arthroplasty: a qualitative and systematic review of the literature. *J Bone Joint Surg [Am]* 2004; 86: 963-74.

[2]Kurtz S, Ong K, Lau E, et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg [Am]* 2007; 89:780

[3] National Joint Registry. www.njrcentre.org.uk. Accessed November 2009

[4]Gidwani S, Fairbank A. The orthopaedic approach to managing osteoarthritis of the knee. *BMJ* 2004; 329:1220-4

[5]O'Brien S, Bennett D, Dora. E, Beaverland D. Comparison of Hip and Knee Arthroplasty Outcomes at Early and Intermediate Follow-up. *Orthopaedics* 2009; 32: 168

[6]Ritter M, Albohm M, Keating E, Faris PM, Meding JB. Comparitive outcomes of total joint arthroplasty. *J Arthroplasty* 1995; 10: 737-41

[7]Norman-Taylor F, Palmer C, Villar R. Quality-of-Life Improvement Compared after Hip and Knee Replacement. *J Bone Joint Surg [Br]* 1996; 78: 74-7

[8]Bachmeier C, March L, Cross M, Lapsley HM, Tribe KL, Courtenay BG, et al. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthritis and Cartilage* 2001; 9: 137-146

[9]Benroth R, Gawand S. Patient-reported health status in total joint replacement. *J Arthroplasty* 1999; 14: 576-80

[10]Bourne RB, Chesworth B, Davis A, et al. Comparing patient outcomes after THA and TKA. Is there a difference. *Clin Orthop Relat Res* 2010;468:542-546

[11]Wylde V, Blom A, Whitehouse SL, Taylor AH, Pattinson GT, Bannister GC. Patient reported outcomes after total hip and knee arthroplasty – comparison of mid-term results. *J Arthroplasty* 2009; 24: 210-6

[12]Dawson J, Doll H, Fitzpatrick R, et al. Letter to the Editor. *J Arthroplasty* 2009; 24: 827

[13]Hand D.J. and Taylor C.C. Multivariate analysis of variance and repeated measures: a practical guide for behavioural scientists, Chapman & Hall. 1987

[14]Ahamed N, Sledge CB, Dalroy L, et al. *Self administered patient satisfaction scale for joint replacement arthroplasty.* Read at the Annual Meeting of the Canadian orthopaedic association 1997 Hamilton ON.

[15]Murray DW, Britton AR, Bulstrode CJK. Loss to follow-up matters. *J Bone Joint Surg Br* 1997;79:254-7

[16]Ware JE, Kosinski M, Keller SD. *SF-12®: How to score the Sf-12® Physical and Mental Health Summary Scales*. Lincoln, RI: QualityMetric Incorporated, Third Edition, 1998.

[17]Fitzmaurice, G. A Conundrum in the Analysis of Change *Nutrition* 2001; 17:360-361

[18]Box, G.E.P. Statistics as a catalyst to learning by scientific method – a discussion. *Journal of Quality Technology*, 1999; 31(1): 16–29

[19]Davies AP. Rating systems for total knee replacement. *Knee* 2002; 9: 261-6

[20]Garratt S, Breakley S, Gillespie JW. Patient assessed health instruments for the knee: a structured review. *Rheumatology* 2004; 43: 1414-23

[21]Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg [Br]* 2007; 89: 1010-4

[22]Hutton JL, Stanghellini E. Modelling bounded health scores with censored skew-normal distributions. *Stat Med.* 2011 Feb 20;30(4):368-76

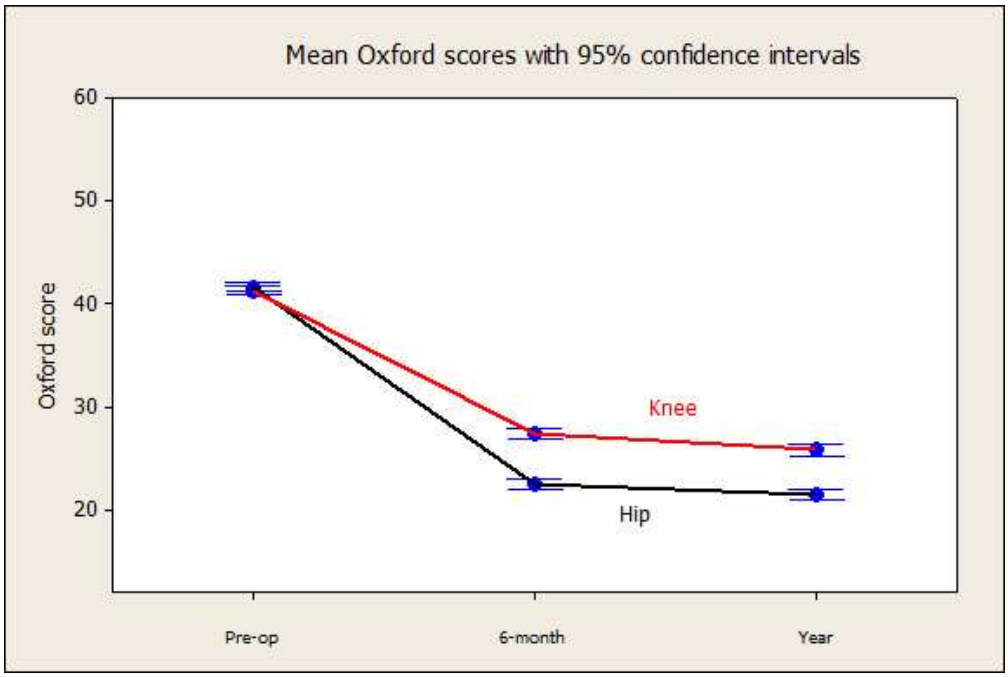


figure 1 - Mean Oxford Scores
206x137mm (71 x 71 DPI)

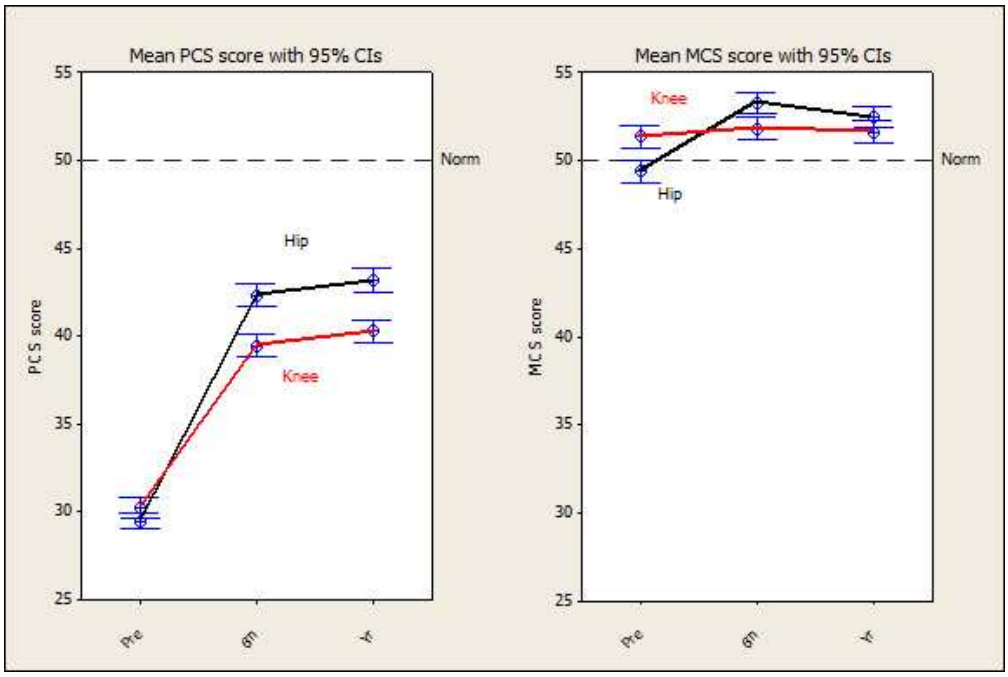


Figure 2 - Mean SF12 Scores
206x137mm (71 x 71 DPI)

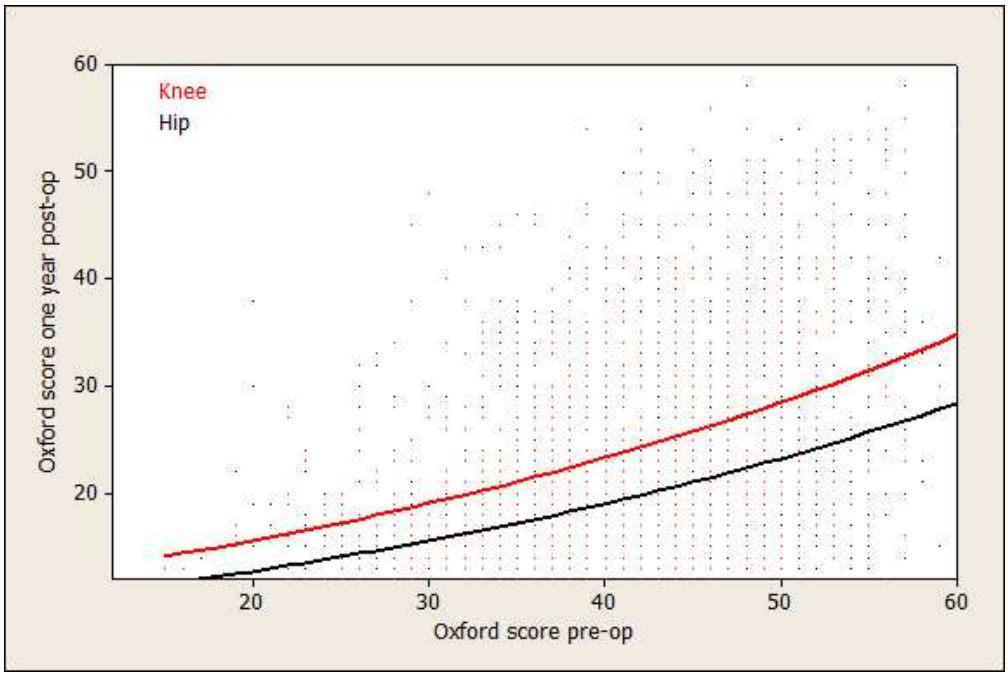


Figure 3 - Scatter plot with ANCOVA model for Oxford Score 1 year post-op
206x137mm (71 x 71 DPI)

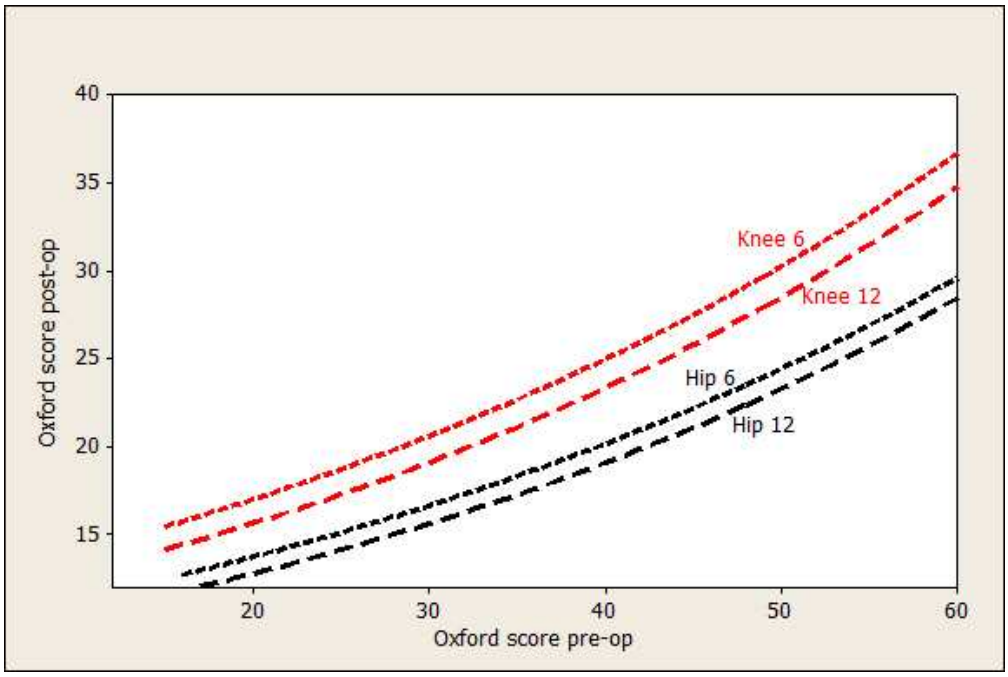


figure 4 - ANCOVA models for Oxford Score at 6 months and 1 year post-op
206x137mm (71 x 71 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	3,4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	3
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	1,4
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	2
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	4
		(*d*) If applicable, explain how loss to follow-up was addressed	4
		(*e*) Describe any sensitivity analyses	
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	2, 4
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	4
		(b) Indicate number of participants with missing data for each variable of interest	2
		(c) Summarise follow-up time (eg, average and total amount)	2
Outcome data	15*	Report numbers of outcome events or summary measures over time	5, 6
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	5, 6
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	7, 8, 9
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8, 9
Generalisability	21	Discuss the generalisability (external validity) of the study results	8, 9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	9

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Appendix B: Conference abstracts

Published abstracts

12th Combined Meeting of the Orthopaedic Associations, Glasgow, 2010

Hamilton DF, Simpson AHRW, Gaston P. *Patient Reported Outcome Measures (PROMS) do not fully represent physical function following total knee arthroplasty.*

Abstract pending publication in JBJS [Br]

Annual meeting of the Physiotherapy Research Society, Middlesburgh, 2010

Hamilton DF, Gaston P, Simpson AHRW. *Assessing outcome post TKA: a model of outcome assessment.* Proceedings of the Physiotherapy Research Society, 2010

Available at: www.prs-uk.org/abstracts/view.php?item=49

British Orthopaedic Research Society, Cardiff, 2010

Hamilton DF, Gaston P, Simpson AHRW. *Muscle satellite cell number influences post operative recovery following total knee replacemen*t.

Abstract pending publication in JBJS [Br]

56th Annual meeting of the Orthopaedic Research Society, New Orleans, 2010

Hamilton DF, Simpson AHRW, Gaston P. *Patient report of outcome differs to objective assessment of physical function following total knee arthroplasty*. Transactions of the Orthopaedic Research Society Vol.35, NewOrleans 2010

Available at: www.ors.org/web/Transactions/56/0410.pdf

Annual meeting of the Physiotherapy Research Society, Glasgow, 2009

Hamilton DF, Simpson AHRW, Gaston P. *Patient Reported Outcome differs to assessment of physical function following total knee arthroplasty*. Proceedings of the Physiotherapy Research Society, 2009

Available at: www.prs-uk.org/abstracts/view.php?item=107

Accepted abstract

3rd joint meeting of the Bone Research Society and British Orthopaedic Research Society, Cambridge, 2011

Hamilton DF, Gaston P, Simpson AHRW. *Single radius of curvature implant design enhances power output following total knee arthroplasty.*

Free paper

PATIENT REPORTED OUTCOME MEASURES (PROMS) DO NOT FULLY REPRESENT PHYSICAL FUNCTION FOLLOWING TOTAL KNEE ARTHROPLASTY

D Hamilton[1, 2], AHRW Simpson[1, 2], P Gaston[2]

[1]University of Edinburgh, Edinburgh, UK,
[2]Royal Infirmary of Edinburgh, Edinburgh, UK

Introduction

Most studies now use Patient Reported Outcome Measures (PROMS) as the preferred and only method for assessing 'functional outcome' following surgery. It is assumed that these questionnaires accurately reflect the patient's pain and physical function. We hypothesised that comprehensive functional examination would therefore correlate strongly with PROMS following total knee arthroplasty (TKA).

Methods

We prospectively assessed the function of 100 consecutive knee replacement patients, pre-operatively, then at 8, 26 and 52 weeks post-operatively. PROMS employed were the Oxford Knee Score (OKS), and the Short Form-36. Additionally, leg strength (Leg Extensor Power Rig TM), a validated battery of timed functional tasks (Aggregated Locomotor Function, ALF), pain scores (numerical rating scale), and range of motion (hand held goniometry) were also assessed. Statistical analysis was performed using the Minitab version 15 software. Level of significance was set as $p = <0.05$.

Results

7 patients were lost to follow-up leaving 93 data sets available for analysis. All of the individual outcome measures showed statistically significant improvement between each of the assessment periods (paired sample t-test, $p = <0.05$). Strong correlation was observed between the two PROMS ($r = 0.74$). Comparatively small correlations were identified between the PROMS and physical assessments ($r = < 0.41$). The pain scores correlated modestly well with the OKS ($r = 0.61$) but comparatively poorly with the physical assessments ($r = < 0.37$).

Discussion

PROMS correlate well with pain, but less well with function following TKA. These results suggest that PROMS may not fully represent actual physical function, but rather the patient's perception of their function. We propose the use of a hierarchical assessment framework to assess patient function. With the advancement of PROMS as the prime outcome measure of most studies, it is important to acknowledge that the assessment provided may be incomplete, and though a useful means of assessing large cohorts, their limitation in assessing functional outcome should be recognised.

Physiotherapy Research Society, Middlesburgh, 2010

Free paper

ASSESSING OUTCOME POST TKA: A MODEL OF OUTCOME ASSESSMENT

D Hamilton (1), AHRW Simpson (1), P Gaston (2)

1 University of Edinburgh, Edinburgh, UK,
2 Royal Infirmary of Edinburgh, Edinburgh, UK

Introduction
Most studies now use Patient Reported Outcome Measures (PROMS) as the preferred and only method for assessing 'functional outcome' following surgery. We hypothesised that comprehensive functional examination would therefore correlate strongly with PROMS following total knee arthroplasty (TKA).

Methods
We prospectively assessed the function of 100 consecutive knee replacement patients, pre-operatively, then at 8, 26 and 52 weeks post-operatively. PROMS employed were the Oxford Knee Score (OKS), and the Short Form-12. Leg strength (Leg Extensor Power RigTM), a validated battery of timed functional tasks (Aggregated Locomotor Function) and pain scores (numerical rating scale) were also assessed. Significance was set as $p = <0.05$.

Results
93 data sets were available for analysis. All the individual outcome measures showed significant improvement between each assessment period (paired sample t-test, $p = <0.05$). Strong correlation was observed between the two PROMS ($r = 0.74$). Comparatively small correlations were identified between the PROMS and physical assessments ($r = < 0.41$). The pain scores correlated comparatively well with the OKS ($r = 0.61$) but comparatively poorly with the physical assessments ($r = < 0.37$).

Discussion
PROMS correlate well with pain, but less well with function following TKA. This suggests that PROMS may not fully represent actual physical function, but perhaps the perception of function. With the advancement of PROMS as the prime outcome measure of most studies, it is important to acknowledge that the assessment may be incomplete, and though a useful assessment of large cohorts, their limitation in assessing functional outcome should be recognised.

Support
This work forms part of a PhD award supported by an MRC doctoral training scholarship and Stryker UK

British Orthopaedic Research Society, Cardiff, 2010

Free Paper

MUSCLE SATELLITE CELL NUMBER INFLUENCES POST OPERATIVE RECOVERY FOLLOWING TOTAL KNEE REPLACEMENT

D.F Hamilton[1], P Gaston[2], A.H.R.W Simpson[1,2]

[1]Orthopaedic Engineering Collaboration, University of Edinburgh, Edinburgh, UK
[2]Dept of Orthopaedics, NHS Lothian

Introduction

Muscle recovery after Total Knee Replacement (TKR) is variable. Satellite cells are undifferentiated myogenic precursors considered to be muscle stem cells. We hypothesised that the recovery of muscle strength following knee replacement in a given patient would be influenced by the underlying number of satellite cells in that patient.

Methods

20 patients undergoing TKR were recruited from the waiting list of a single consultant. A muscle biopsy was taken at the time of surgery from the distal quadriceps. This was fixed in paraffin wax, and sections obtained. Satellite cells were identified with a primary mouse antibody for Pax7 - a cytoplasmic protein marker - and an immunofluorescent goat anti-mouse secondary. Slides were counterstained with DAPI to stain the myonuclei. The positive staining index (PSI) was calculated (number of satellite cells/total number of myonuclei x 100). Recovery of muscle (quadriceps) strength was assessed using the leg extensor power-rig (LegRig), pre-operatively, at 6 and 26 weeks post-operatively. Statistical analysis was performed on the Minitab version 15 software, level of significance was set as $p = 0.05$

Results

3 patients were unable to provide follow-up data. The number of satellite cells varied (PSI 3.07 to 11.35). Improvement in muscle power varied (0 to 70 W) between the 6 and 26 weeks assessment periods. This reflected a 0 to 60% improvement in the individual's strength to bodyweight ratio. The improvement in muscle power correlated with the satellite cell numbers (determined at the time of surgery). This was true for both absolute improvement in wattage generated, $r = 0.54$ $p= 0.038$ and improvement in strength relative to body weight $r = 0.47$ $p = 0.06$. Linear regression analysis demonstrated that the relative satellite cell number accounted for 30% of the improvement in muscle.

Discussion

We have for the first time demonstrated that the magnitude of improvement in muscle strength following TKR may be influenced by the patient¹s underlying pool of satellite cells, with up to 30% of the variation of improvement in our cohort attributable to the satellite cell pool.

Orthopaedic Research Society, New Orleans, 2010

Patient Report of Outcome Differs to Objective Assessment of Physical Function Following Total Knee Arthroplasty

+[1]Hamilton, D F; [2]Simpson, A H R W; [2]Gaston, P
+[1]University of Edinburgh, Edinburgh, UK, [2]Royal Infirmary of Edinburgh, Edinburgh, UK

Introduction
End-stage osteoarthritis is characterised by pain and reduced physical function, for which total knee arthroplasty (TKA) is recognised to be a highly effective treatment. It is important to quantify improvement following surgical intervention, and increasingly Patient Reported Outcome Measures (PROMS) are the preferred method of assessing pain and physical function following TKA, Most studies now utilise these self-report measures alone to report patient functional outcome. It is assumed that these self-report outcome questionnaires accurately reflect the patient's pain and physical function. Limited evidence is now emerging however that this may not be the case. We hypothesised that a physical examination of patient function would correlate strongly with patient self-report questionnaires following TKA.

Methods
Approval was granted by the local ethical research committee and 100 consecutive patients with knee osteoarthritis listed for TKA at the investigating hospital were recruited to the study with informed consent. Assessment was carried out pre-operatively and at 8 weeks and 26 weeks post-operation. Patient report questionnaires were completed independently and handed to the researcher in a sealed envelope prior to the physical assessments. The self-report outcome questionnaires used were the Oxford Knee Score (OKS) which assesses pain and function following TKA as a single score, and the Short Form-36 which assesses physical function (PCS) and mental function (MCS) separately. Additionally, direct measurement of leg strength was assessed (Leg Extensor Power Rig TM), a validated performance battery of timed functional tasks was performed (Automated Locomotor Function, ALF), pain scores were recorded by numerical rating scale, and range of motion was assessed by hand held goniometry. All assessments were carried out by the same researcher. Statistical analysis was performed using the Minitab version 15 software. Level of significance was set as $p = <0.05$.

Results
Of the 100 recruits, 93 complete data sets were available for analysis. All of the individual outcome measures showed statistically significant improvement between each of the assessment periods (paired sample t-test, $p = <0.05$). Scatter plots were created for the interaction between the outcome variables. Normal distribution was found in all cases; therefore the Pearson correlation coefficient was calculated. Strong correlation was observed between the two separate patient report measures ($r = 0.74$, $p = <0.001$). Lesser correlation was observed between the separate physical assessments, ALF and leg strength ($r = 0.47$, $p = 0.001$). Comparatively small correlations were identified between the patient report and physical assessments. The OKS correlated with the ALF, ($r = 0.37$, $p = < 0.001$) and with leg strength ($r = 0.37$, $p = 0.001$), further the PCS correlated with ALF ($r = 0.41$, $p = < 0.001$) and with leg strength ($r = 0.18$, $p = 0.15$). The pain scores correlated moderately well with the OKS ($r = 0.61$, $p = < 0.001$) and comparatively poorly with the other measures; with PCS, $r = 0.38$ ($p = 0.002$), with ALF, $r = 0.37$ ($p = < 0.001$), with leg strength, $r = 0.15$ ($p = 0.154$). Range of motion formed small correlations with all of the other outcome assessments; with OKS, $r = 0.31$ ($p = 0.002$), with PCS, $r = 0.26$ ($p = 0.04$), with ALF, $r = 0.34$ ($p = 0.001$), with leg strength, $r = 0.23$ ($p = 0.003$), with pain score, $r = 0.32$ ($p = 0.002$).

Discussion
As expected, significant improvements in outcome scores were detected across all our outcome measures and between all the assessment periods, charting the improvement that accompanies recovery following TKA surgery. Among our outcome assessments, interestingly, strong correlation was found between the two separate patient questionnaires, but not between the questionnaire scores and direct physical assessment. It is notoriously difficult to interpret correlation coefficients. It has been suggested that any criteria set are in some way arbitrary as interpretation depends upon the context of the investigation[2]. In this context, it is reassuring that the expected strong correlation between the two separate PROMS was found. This puts into perspective the poor-to-modest correlations found between the patient report scores and the physical assessments that these scores are supposed to represent. Patient outcome following TKA is of complex multifactorial nature. That the leg strength and functional assessments correlate only modestly suggests that these associated measures assess different aspects of patient function. Of further interest is the role of range of motion. Our results suggest a small but significant interaction between knee flexion and patient outcome, with less flexion being associated with poorer outcome scores across our assessments. The interaction of these factors reinforces this multifactorial nature of functional outcome. Our results further confirm the known link between patient postoperative pain, and functional outcome. It is interesting however, that the patient report of pain correlated strongly with the patient report questionnaires (that include specific questions addressing pain) but substantially less well with the physical assessments. This finding, while acknowledging the good agreement of pain scores, further questions the ability of the PROMS to reflect patient function. It may be, as has been suggested[1], that the self-report measures represent the experience of the patient when performing an activity, rather than the patient's ability to actually perform that activity. These results suggest that self-report outcome measures may not fully represent actual physical function following TKA. With the advancement of PROMS in the surgical literature, it is important to recognise that the assessment they provide may be incomplete, and that the practice of reporting functional outcome based solely on patient report methods should be questioned. A larger longitudinal study would be beneficial to fully assess this issue.

Acknowledgments
This work forms part of a PhD award funded jointly by the Medical Research Council (UK) Doctoral Training Scheme and by Stryker UK.

Physiotherapy Research Society, Glasgow, 2009

Free paper

PATIENT REPORTED OUTCOME DIFFERS TO ASSESSMENT OF PHYSICAL FUNCTION FOLLOWING TOTAL KNEE ARTHROPLASTY

DF. Hamilton[1], AHRW. Simpson[1,2], P. Gaston[2]

[1] Edinburgh Orthopaedic Engineering Collaboration, University of Edinburgh, UK
[2] Royal Infirmary of Edinburgh, UK

Purpose
To examine the relationship between physical outcome and patient report of outcome following total knee arthroplasty (TKA).

Relevance
Increasingly Patient Reported Outcome Measures (PROMS) are used to assess function following TKA. It is assumed that these measures accurately reflect the patient's pain and physical function, however, evidence is emerging that this may incorrect.

Methods
Approval was granted by the Local Research Ethics Committee, and 26 consecutive patients listed for TKA were assessed pre TKA, and at 8 and 26 weeks post-operation. The Oxford Knee Score (OKS) and the physical function score (PCS) of the Short Form-36 were utilised as self report outcomes. Direct measurement of leg strength was assessed (Leg extensor Power Rig TM). Timed assessment of functional tasks was performed (Automated Locomotor Function, ALF). Pain scores were recorded by numerical rating scale.

Results
All the individual measures showed statistically significant improvement between each of the assessment periods (paired sample t-test, $p = <0.05$). The PROMS however were found to correlate poorly to the physical assessments. The OKS correlated poorly with the ALF, ($r = < 0.5$) and with leg strength ($r = < 0.3$).The PCS correlated poorly with ALF ($r = < 0.2$) and with leg strength ($r = 0.34$).

Conclusions
These results suggest that PROMS do not fully represent actual physical function following TKA. With the advancement of PROMS in the literature, it is important to recognise that the assessment they provide may be incomplete. A larger study would be beneficial to assess this issue fully.

Support
This work forms part of a PhD award with is jointly funded by the Medical Research Council and Stryker UK

Additional peer reviewed presentations

American Travelling Hip Fellows, Edinburgh, 2010

Invited speaker: Comparative THA and TKA outcome

Association of Orthopaedic Physiotherapists: national course, Edinburgh, 2009

Invited Lecture: Assessing TKA outcome

British Orthopaedic Research Society, Newcastle, 2009

Poster: TKR Outcome: PROMS not consistent with direct function assessment

ABC Travelling Fellows, Edinburgh, 2009

Invited speaker: Outcome assessments – PROMS are not the whole story

SICOT, Pattaya, 2009

Poster: Differences exist between PROMS and direct physical assessment measures

Appendix C: Statistical output (Chapter 3)

Oxford Knee Score

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre Ox	1	41.36	7.220	95
	2	40.32	7.915	79
	Total	40.89	7.539	174
12m Ox	1	21.35	7.679	95
	2	23.19	8.290	79
	Total	22.18	7.992	174

Group * Oxford

				95% Confidence Interval	
Group	oxford	Mean	Std. Error	Lower Bound	Upper Bound
1	1	41.358	.774	39.830	42.886
	2	21.347	.817	19.735	22.960
2	1	40.316	.849	38.641	41.992
	2	23.190	.896	21.422	24.958

Tests of Within-Subjects Contrasts

Source	oxford	Type III Sum of Squares	df	Mean Square	F	Sig.
Oxford	Linear	29743.196	1	29743.196	709.462	.000
Oxford * Group	Linear	179.368	1	179.368	4.278	.040
Error (oxford)	Linear	7210.862	172	41.924		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Wilcoxon Signed Ranks Test

Test Statistics [b]

Group		6w Ox – pre Ox	6m Ox – 6w Ox	12m Ox – 6m Ox
1	Z	-7.495[a]	-7.211[a]	-5.821[a]
	Asymp. Sig. (2-tailed)	.000	.000	.000
2	Z	-5.401[a]	-6.885[a]	-3.639[a]
	Asymp. Sig. (2-tailed)	.000	.000	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Mann-Whitney U-test

Test Statistics [a]

	Pre Ox	6w Ox	6m Ox	12m Ox
Mann-Whitney U	3525.000	3396.500	3550.500	3331.000
Wilcoxon W	6685.000	7674.500	6325.500	8381.000
Z	-.801	-.172	-.120	-2.180
Asymp. Sig. (2-tailed)	.423	.863	.904	.029

a. Grouping Variable: Group

Range of motion – Flexion

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre Flex	1	104.85	14.762	100
	2	104.52	14.543	82
	Total	104.70	14.624	182
12m Flex	1	110.00	10.863	100
	2	103.88	12.025	82
	Total	107.24	11.772	182

Group * flexion

				95% Confidence Interval	
Group	flexion	Mean	Std. Error	Lower Bound	Upper Bound
1	1	104.850	1.466	101.956	107.744
	2	110.000	1.140	107.750	112.250
2	1	104.524	1.619	101.329	107.720
	2	103.878	1.259	101.394	106.362

Tests of Within-Subjects Contrasts

Source	flexion	Type III Sum of Squares	Df	Mean Square	F	Sig.
flexion	Linear	456.923	1	456.923	5.082	.025
flexion * Group	Linear	756.868	1	756.868	8.418	.004
Error (flexion)	Linear	16184.747	180	89.915		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Paired Samples Test

Group		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper			
1	Flex pre – Flex 6w	7.323	14.714	1.502	4.34	10.30	4.876	95	.000
	Flex 6w – Flex 6m	-8.928	9.400	.954	-10.82	-7.03	-9.35	96	.000
	Flex 6m– Flex 12m	-3.673	6.391	.636	-4.94	-2.41	-5.78	100	.000
2	preFlex - 6wFlex	10.200	13.032	1.46	7.30	13.1	7.01	79	.000
	6wFlex - 6mFlex	-7.532	8.001	.91	-9.35	-5.72	-8.26	76	.000
	6mFlex - 12mFlex	-2.872	6.925	.78	-4.43	-1.31	-3.66	77	.000

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
12m Flex	assumed	.012	.913	3.660	181	.000	6.191	1.692	2.853	9.529
	not assumed			3.620	164.90	.000	6.191	1.710	2.814	9.568
6m Flex	assumed	.005	.944	2.661	177	.009	4.870	1.830	1.258	8.483
	not assumed			2.637	159.84	.009	4.870	1.847	1.223	8.517
6w Flex	assumed	.520	.472	1.701	175	.091	3.380	1.987	-.541	7.301
	not assumed			1.705	170.08	.090	3.380	1.982	-.532	7.292
Pre Flex	assumed	.003	.956	.149	180	.882	.326	2.185	-3.985	4.636
	not assumed			.149	174.04	.882	.326	2.181	-3.980	4.631

Range of motion – Extension

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre Ext	1	3.05	5.239	100
	2	3.27	5.698	82
	Total	3.15	5.436	182
12m Ext	1	.30	1.210	100
	2	1.16	3.320	82
	Total	.69	2.433	182

Group * extension

Group	extension	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	3.050	.545	1.975	4.125
	2	.300	.240	-.174	.774
2	1	3.268	.602	2.081	4.456
	2	1.159	.265	.635	1.682

Tests of Within-Subjects Contrasts

Source	extension	Type III Sum of Squares	df	Mean Square	F	Sig.
extension	Linear	532.036	1	532.036	38.703	.000
extension * Group	Linear	9.234	1	9.234	.672	.414
Error (extension)	Linear	2474.381	180	13.747		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Paired Samples Test

Group		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper			
1	Ext pre – Ext 6w	1.13	5.79	.59	-.05	2.30	1.90	95	.060
	Ext 6w – Ext 6m	1.22	3.00	.31	.61	1.82	3.99	96	.000
	Ext 6m – Ext 12m	.54	1.56	.15	.23	.84	3.45	100	.001
2	Ext pre – Ext 6w	1.29	5.37	.60	.09	2.48	2.14	79	.035
	Ext 6w – Ext 6m	1.09	3.03	.35	.40	1.78	3.16	76	.002
	Ext 6m – Ext 12m	.09	1.58	.18	-.27	.45	.50	77	.617

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
Ext pre	assumed	.510	.476	-.269	180	.788	-.218	.812	-1.821	1.384
	not assumed			-.267	166.68	.790	-.218	.819	-1.835	1.398
Ext 6w	assumed	.271	.603	.134	175	.894	.081	.605	-1.113	1.275
	not assumed			.134	167.96	.894	.081	.605	-1.114	1.276
Ext 6m	assumed	1.120	.291	-.445	177	.657	-.155	.350	-.846	.535
	not assumed			-.432	144.18	.667	-.155	.360	-.867	.556
Ext 12m	assumed	22.784	.000	-2.42	181	.017	-.862	.356	-1.564	-.159
	not assumed			-2.23	98.35	.028	-.862	.386	-1.627	-.096

Timed functional assessment

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre ALF	1	36.1769	15.63527	101
	2	34.3961	12.32829	82
	Total	35.3790	14.23906	183
12m ALF	1	25.5044	6.83123	101
	2	25.7340	6.34935	82
	Total	25.6073	6.60270	183

Group * ALF

				95% Confidence Interval	
Group	ALF	Mean	Std. Error	Lower Bound	Upper Bound
1	1	36.177	1.418	33.379	38.975
	2	25.504	.659	24.205	26.804
2	1	34.396	1.574	31.291	37.501
	2	25.734	.731	24.292	27.176

Tests of Within-Subjects Contrasts

Source	ALF	Type III Sum of Squares	df	Mean Square	F	Sig.
ALF	Linear	8459.149	1	8459.149	127.036	.000
ALF * Group	Linear	91.467	1	91.467	1.374	.243
Error (ALF)	Linear	12052.493	181	66.588		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Paired Samples Test

Group		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper			
1	ALF pre – ALF 6w	4.64	11.82	1.19	2.27	7.01	3.885	97	.000
	ALF 6w – ALF 6m	4.93	6.66	.67	3.59	6.26	7.317	97	.000
	ALF 6m – ALF 12m	1.10	3.99	.40	.31	1.89	2.761	100	.007
2	ALF pre – ALF 6w	2.82	10.09	1.13	.57	5.06	2.497	79	.015
	ALF 6w – ALF 6m	5.50	7.74	.88	3.74	7.26	6.233	76	.000
	ALF 6m – ALF 12m	.84	3.99	.45	-.065	1.74	1.848	77	.068

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
ALF pre	assumed	1.124	.290	.841	181	.402	1.78	2.12	-2.40	5.96
	not assumed			.861	180.86	.390	1.78	2.07	-2.30	5.86
ALF 6w	assumed	.161	.689	-.057	176	.955	-.10	1.74	-3.53	3.33
	not assumed			-.057	173.21	.955	-.10	1.73	-3.50	3.31
ALF 6m	assumed	.469	.494	.099	177	.921	.12	1.18	-2.20	2.44
	not assumed			.100	169.86	.920	.12	1.17	-2.19	2.42
ALF 12m	assumed	.221	.639	-.233	181	.816	-.23	.98	-2.17	1.712
	not assumed			-.235	177.68	.814	-.23	.98	-2.16	1.70

Lower limb power output

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre LR Max	1	42.20	37.809	99
	2	47.11	44.298	80
	Total	44.40	40.792	179
12m LR Max	1	87.22	48.380	99
	2	77.59	46.561	80
	Total	82.92	47.686	179

Group * Legrig

				95% Confidence Interval	
Group	Legrig	Mean	Std. Error	Lower Bound	Upper Bound
1	1	42.202	4.104	34.103	50.301
	2	87.222	4.782	77.786	96.659
2	1	47.113	4.565	38.103	56.122
	2	77.586	5.319	67.089	88.084

Tests of Within-Subjects Contrasts

Source	Legrig	Type III Sum of Squares	df	Mean Square	F	Sig.
Legrig	Linear	126085.887	1	126085.887	239.985	.000
Legrig * Group	Linear	4681.190	1	4681.190	8.910	.003
Error (Legrig)	Linear	92994.007	177	525.390		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Paired Samples Test

Group		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper			
1	LR pre – LR 6w	-11.47	22.49	2.30	-16.03	-6.91	-5.00	95	.000
	LR 6w -LR 6m	-20.07	21.05	2.13	-24.29	-15.85	-9.44	97	.000
	LR 6m – LR 12m	-12.70	21.93	2.182	-17.03	-8.37	-5.82	100	.000
2	LR pre – LR 6w	-1.86	30.26	3.43	-8.68	4.96	-.54	77	.589
	LR 6w – LR 6m	-25.47	26.75	3.05	-31.54	-19.40	-8.35	76	.000
	LR 6m – LR 12m	-4.51	24.48	2.77	-10.03	1.01	-1.62	77	.108

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
Pre LR	assumed	1.101	.296	-.800	177	.425	-4.91	6.139	-17.03	7.204
	not assumed			-.787	155.84	.433	-4.91	6.242	-17.24	7.420
6w LR	assumed	3.338	.069	.876	176	.382	4.745	5.41	-5.940	15.43
	not assumed			.891	175.77	.374	4.745	5.32	-5.761	15.25
6m LR	assumed	.001	.974	.152	177	.879	.972	6.38	-11.61	13.55
	not assumed			.151	157.72	.880	.972	6.45	-11.77	13.71
12m LR	assumed	1.681	.196	1.435	181	.153	10.062	7.01	-3.777	23.90
	not assumed			1.441	176.18	.151	10.062	6.98	-3.715	23.84

Proportional lower limb power output

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
LR Prop Pre	1	50.4060	32.04205	98
	2	50.4655	36.67287	80
	Total	50.4655	36.67287	80
LR Prop 12M	1	116.5360	35.91522	98
	2	93.3504	35.77775	80
	Total	93.3504	35.77775	80

Group * Legrig proportional

Group	Legrig prop	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	50.406	3.455	43.588	57.224
	2	116.536	3.622	109.388	123.684
2	1	50.466	3.823	42.920	58.011
	2	93.350	4.009	85.439	101.261

Tests of Within-Subjects Contrasts

Source	Legrig prop	Type III Sum of Squares	df	Mean Square	F	Sig.
Legrig prop	Linear	261720.588	1	261720.588	249.087	.000
Legrig prop * Group	Linear	11899.586	1	11899.586	11.325	.001
Error (leg rig prop)	Linear	184926.290	176	1050.718		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Paired Samples Test

Group		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper			
1	LR Pre – LR 6w	-17.06	36.78	3.77	-24.56	-9.57	-4.52	94	.000
	LR 6w – LR 6m	-34.76	44.68	4.53	-43.76	-25.75	-7.66	96	.000
	LR 6m – LR 12M	-14.07	42.04	4.20	-22.41	-5.73	-3.35	99	.001
2	LR Pre – LR 6w	-6.92	37.61	4.26	-15.40	1.55	-1.63	77	.108
	LR 6w – LR 6m	-30.87	28.58	3.26	-37.35	-24.38	-9.48	76	.000
	LR 6m – LR 12M	-6.03	28.04	3.18	-12.35	.28	-1.90	77	.061

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
LR Pre	assumed	2.354	.127	-.012	176	.991	-.06	5.15	-10.22	10.10
	not assumed			-.011	158.13	.991	-.06	5.22	-10.37	10.25
LR 6w	assumed	.029	.865	2.277	175	.024	10.6	4.67	1.42	19.84
	not assumed			2.317	174.98	.022	10.6	4.59	1.58	19.68
LR 6m	assumed	3.941	.049	2.081	176	.039	13.37	6.42	.69	26.04
	not assumed			2.207	165.93	.029	13.37	6.06	1.41	25.33
LR 12M	assumed	.007	.934	4.427	180	.000	23.42	5.29	12.98	33.86
	not assumed			4.428	173.33	.000	23.42	5.29	12.98	33.86

Reported average pain score

1 year outcome ANOVA output

Statistical analysis (SPSS output):

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre P Ave	1	5.24	1.511	101
	2	5.54	1.557	82
	Total	5.37	1.535	183
12m P Ave	1	.83	1.484	101
	2	1.16	1.567	82
	Total	.98	1.526	183

Group * Pain ave

Group	Pain ave	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	5.238	.152	4.937	5.538
	2	.832	.151	.533	1.130
2	1	5.537	.169	5.203	5.870
	2	1.159	.168	.827	1.490

Tests of Within-Subjects Contrasts

Source	Pain ave	Type III Sum of Squares	df	Mean Square	F	Sig.
Pain ave	Linear	1745.974	1	1745.974	849.934	.000
Pain ave * Group	Linear	.018	1	.018	.009	.926
Error (Pain ave)	Linear	371.818	181	2.054		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Wilcoxon Signed Ranks Test

Test Statistics [b]

Group		6w P Ave – pre P Ave	6m P Ave – 6w P Ave	12m P Ave - 6m P Ave
1	Z	-7.513[a]	-6.112[a]	-3.795[a]
	Asymp. Sig. (2-tailed)	.000	.000	.000
2	Z	-6.116[a]	-5.529[a]	-2.944[a]
	Asymp. Sig. (2-tailed)	.000	.000	.003

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Mann-Whitney U-test

Test Statistics [a]

	Pre P Ave	6w P Ave	6m P Ave	12m P Ave
Mann-Whitney U	3689.000	3402.500	3493.000	3599.500
Wilcoxon W	8840.000	8253.500	8644.000	8750.500
Z	-1.302	-1.535	-1.370	-1.741
Asymp. Sig. (2-tailed)	.193	.125	.171	.082

a. Grouping Variable: Group

Reported maximal pain score

1 year outcome ANOVA output

Statistical analysis (SPSS) output:

Descriptive Statistics

	Group	Mean	Std. Deviation	N
Pre P Max	1	8.29	1.410	101
	2	8.24	1.520	82
	Total	8.27	1.456	183
12m P Max	1	1.90	2.452	101
	2	2.90	2.618	82
	Total	2.35	2.569	183

Group * pain max

Group	Pain max	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	8.287	.145	8.000	8.574
	2	1.901	.251	1.405	2.397
2	1	8.244	.161	7.926	8.562
	2	2.902	.279	2.352	3.453

Tests of Within-Subjects Contrasts

Source	Pain max	Type III Sum of Squares	df	Mean Square	F	Sig.
Pain max	Linear	3112.236	1	3112.236	937.000	.000
Pain max * Group	Linear	24.695	1	24.695	7.435	.007
Error (pain max)	Linear	601.190	181	3.321		

Longitudinal between assessment and between group output

Statistical analysis (SPSS) output:

Wilcoxon Signed Ranks Test

Test Statistics[b]

Group		6w P Max – pre p Max	6m P Max – 6w P Max	12m P Max - 6m P Max
1	Z	-7.558[a]	-6.760[a]	-3.884[a]
	Asymp. Sig. (2-tailed)	.000	.000	.000
2	Z	-6.721[a]	-6.218[a]	-1.915[a]
	Asymp. Sig. (2-tailed)	.000	.000	.055

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Mann-Whitney U-test

Test Statistics [a]

	Pre P Max	6w P Max	6m P Max	12m P Max
Mann-Whitney U	4141.000	3485.000	3744.500	3148.500
Wilcoxon W	7544.000	8336.000	8895.500	8299.500
Z	.000	-1.282	-.576	-2.885
Asymp. Sig. (2-tailed)	1.000	.200	.565	.004

a. Grouping Variable: Group

Appendix D: Statistical output (Chapter 4)

Multiple linear regression analysis of predictors of OKS (12 month assessment)

```
The regression equation is:

12mOx = 22.2 - 0.0367 12mFlex + 0.314 12mALF + 1.24 12mPMax + 1.12 12mPAve
        - 0.0317 12mLRMax - 1.12 Gender - 0.0550 Age

182 cases used, 1 cases contain missing values

Predictor      Coef  SE Coef      T      P
Constant     22.236    5.561   4.00  0.000
12mFlex    -0.03668  0.03335  -1.10  0.273
12mALF      0.31418  0.06907   4.55  0.000
12mPMax      1.2354   0.2218   5.57  0.000
12mPAve      1.1157   0.3793   2.94  0.004
12mLRMax   -0.03174  0.01041  -3.05  0.003
Gender      -1.1166   0.8880  -1.26  0.210
Age        -0.05498  0.04801  -1.15  0.254

S = 4.83721   R-Sq = 64.0%   R-Sq(adj) = 62.5%

PRESS = 4557.76   R-Sq(pred) = 59.69%

Analysis of Variance

Source           DF       SS      MS      F      P
Regression        7   7236.4  1033.8  44.18  0.000
Residual Error  174   4071.4    23.4
Total           181  11307.7
```

Stepwise regression model

```
Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is 12mOx on 7 predictors, with N = 182
N (cases with missing observations) = 1 N (all cases) = 183

Step             1      2      3        4
Constant    17.122  7.245  7.776   11.635

12mPMax       2.14   1.88   1.34     1.33
T-Value      12.97  12.42   6.07     6.16
P-Value      0.000  0.000  0.000    0.000

12mALF               0.409  0.391    0.317
T-Value               6.96   6.81     5.02
P-Value              0.000  0.000    0.000

12mPAve                     1.22     1.16
T-Value                     3.30     3.16
P-Value                    0.001    0.002

12mLRMax                           -0.0227
T-Value                              -2.61
P-Value                              0.010

S             5.70   5.07   4.93     4.86
R-Sq         48.30  59.32  61.67    63.09
R-Sq(adj)    48.01  58.86  61.02    62.25
Mallows Cp    71.9   20.6   11.2      6.4
```

Multiple linear regression analysis of predictors of OKS (6 month assessment)

The regression equation is:

6mOx = 22.8 + 0.0153 6mFlex + 0.415 6mALF + 0.988 6mPMax + 0.841 6mPAve
- 0.0322 6mLRMax - 1.19 Sex - 0.152 Age

171 cases used, 12 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	22.823	5.773	3.95	0.000
6mFlex	0.01531	0.03596	0.43	0.671
6mALF	0.41537	0.06699	6.20	0.000
6mPMax	0.9885	0.2805	3.52	0.001
6mPAve	0.8408	0.4343	1.94	0.055
6mLRMax	-0.03216	0.01304	-2.47	0.015
Gender	-1.1914	0.9852	-1.21	0.228
Age	-0.15233	0.05248	-2.90	0.004

S = 5.37420 R-Sq = 59.2% R-Sq(adj) = 57.5%

PRESS = 5272.48 R-Sq(pred) = 54.32%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	6834.01	976.29	33.80	0.000
Residual Error	163	4707.78	28.88		
Total	170	11541.79			

Stepwise regression model

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is 6mOx on 7 predictors, with N = 171
N (cases with missing observations) = 12 N (all cases) = 183

Step	1	2	3	4
Constant	19.224	9.117	17.404	22.417
6mPMax	1.93	1.59	1.47	1.45
T-Value	10.55	9.54	8.72	8.74
P-Value	0.000	0.000	0.000	0.000
6mALF		0.417	0.474	0.405
T-Value		7.34	7.96	6.27
P-Value		0.000	0.000	0.000
Age			-0.138	-0.153
T-Value			-2.72	-3.03
P-Value			0.007	0.003
6mLRMax				-0.029
T-Value				-2.51
P-Value				0.013
S	6.42	5.60	5.50	5.41
R-Sq	39.69	54.35	56.29	57.89
R-Sq(adj)	**39.34**	**53.81**	**55.50**	**56.87**
Mallows Cp	74.0	17.4	11.7	7.3

Multiple linear regression analysis of predictors of OKS (6 week assessment)

The regression equation is:

6wOx = 36.7 - 0.0243 6wFlex + 0.157 6wALF + 0.884 6wPMax + 1.36 6wPAve
- 0.0472 6wLRMax - 2.34 Sex - 0.137 Age

166 cases used, 17 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	36.660	6.703	5.47	0.000
6wFlex	-0.02428	0.04258	-0.57	0.569
6wALF	0.15668	0.05536	2.83	0.005
6wPMax	0.8839	0.3135	2.82	0.005
6wPAve	1.3561	0.3917	3.46	0.001
6wLRMax	-0.04723	0.02011	-2.35	0.020
Gender	-2.338	1.284	-1.82	0.071
Age	-0.13730	0.06168	-2.23	0.027

S = 6.54052 R-Sq = 48.5% R-Sq(adj) = 46.2%

PRESS = 7452.01 R-Sq(pred) = 43.18%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	6356.41	908.06	21.23	0.000
Residual Error	158	6758.99	42.78		
Total	165	13115.40			

Stepwise regression model

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is 6wOx on 7 predictors, with N = 166
N (cases with missing observations) = 17 N (all cases) = 183

Step	1	2	3
Constant	21.10	16.20	16.98
6wPMax	2.10	1.83	1.11
T-Value	9.46	8.22	3.57
P-Value	0.000	0.000	0.000
6wALF		0.202	0.179
T-Value		4.11	3.70
P-Value		0.000	0.000
6wPAve			1.29
T-Value			3.25
P-Value			0.001
S	7.19	6.87	6.67
R-Sq	35.32	41.38	44.98
R-Sq(adj)	**34.92**	**40.66**	**43.96**
Mallows Cp	36.3	19.7	10.7

Multiple linear regression analysis of predictors of OKS (Pre-operative assessment)

The regression equation is:

preOx = 37.7 - 0.0637 preFlex + 0.153 preALF + 0.904 prePMax + 1.08 prePAve
- 0.0341 preLRMax - 1.35 Sex - 0.0751 Age

171 cases used, 12 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	37.731	6.873	5.49	0.000
preFlex	-0.06375	0.03484	-1.83	0.069
preALF	0.15347	0.03816	4.02	0.000
prePMax	0.9043	0.3748	2.41	0.017
prePAve	1.0796	0.3884	2.78	0.006
preLRMax	-0.03409	0.01479	-2.30	0.022
Gender	-1.353	1.111	-1.22	0.225
Age	-0.07509	0.05630	-1.33	0.184

S = 6.02596 R-Sq = 38.6% R-Sq(adj) = 36.0%

PRESS = 6525.65 R-Sq(pred) = 32.36%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	3728.77	532.68	14.67	0.000
Residual Error	163	5918.89	36.31		
Total	170	9647.66			

Stepwise regression model

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is preOx on 7 predictors, with N = 171
N (cases with missing observations) = 12 N (all cases) = 183

Step	1	2	3	4
Constant	29.25	24.19	18.61	28.77
prePAve	2.17	1.88	1.40	1.21
T-Value	6.23	5.77	3.81	3.26
P-Value	0.000	0.000	0.000	0.001
preALF		0.188	0.192	0.169
T-Value		5.60	5.81	4.98
P-Value		0.000	0.000	0.000
prePMax			0.97	0.99
T-Value			2.63	2.72
P-Value			0.009	0.007
preFlex				-0.081
T-Value				-2.36
P-Value				0.020
S	6.81	6.28	6.17	6.08
R-Sq	18.65	31.43	34.16	36.29
R-Sq(adj)	**18.17**	**30.62**	**32.98**	**34.76**
Mallows Cp	49.1	17.2	11.9	8.3

Multiple linear regression analysis of pre-operative predictors of 12 month OKS

The regression equation is:

12mOx = 31.3 - 0.0429 preFlex + 0.0759 preALF + 0.739 prePMax - 0.491 prePAve - 0.0444 preLRMax - 2.41 Sex - 0.0700 Age

177 cases used, 6 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	31.311	8.539	3.67	0.000
preFlex	-0.04294	0.04354	-0.99	0.325
preALF	0.07592	0.04786	1.59	0.115
prePMax	0.7388	0.4700	1.57	0.118
prePAve	-0.4905	0.4612	-1.06	0.289
preLRMax	-0.04444	0.01879	-2.36	0.019
Sex	-2.406	1.405	-1.71	0.089
Age	-0.06998	0.07018	-1.00	0.320

S = 7.66594 R-Sq = 10.6% R-Sq(adj) = 6.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1175.44	167.92	2.86	0.008
Residual Error	169	9931.55	58.77		
Total	176	11106.99			

Stepwise regression model

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is 12mOx on 8 predictors, with N = 170
N (cases with missing observations) = 13 N (all cases) = 183

Step	1
Constant	9.528
preOx	0.313
T-Value	4.01
P-Value	0.000
S	7.68
R-Sq	8.74
R-Sq(adj)	**8.19**
Mallows Cp	2.6

Multiple linear regression analysis of pre-operative predictors of 6 month OKS

The regression equation is:

6mOx = 38.1 - 0.0750 preFlex + 0.132 preALF + 0.400 prePMax - 0.196 prePAve
- 0.0290 preLRMax - 1.41 Sex - 0.125 Age

166 cases used, 17 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	38.097	9.119	4.18	0.000
preFlex	-0.07495	0.04664	-1.61	0.110
preALF	0.13214	0.05027	2.63	0.009
prePMax	0.4004	0.5051	0.79	0.429
prePAve	-0.1963	0.4858	-0.40	0.687
preLRMax	-0.02899	0.02000	-1.45	0.149
Sex	-1.412	1.465	-0.96	0.336
Age	-0.12453	0.07504	-1.66	0.099

S = 7.80571 R-Sq = 14.3% R-Sq(adj) = 10.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1606.79	229.54	3.77	0.001
Residual Error	158	9626.80	60.93		
Total	165	11233.59			

Stepwise regression model

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is 6mOx on 8 predictors, with N = 160
N (cases with missing observations) = 23 N (all cases) = 183

Step	1
Constant	5.970
preOx	0.466
T-Value	5.82
P-Value	0.000
S	7.59
R-Sq	17.66
R-Sq(adj)	**17.14**
Mallows Cp	1.1

Multiple linear regression analysis of pre-operative predictors of 6 week OKS

The regression equation is:

6wOx = 28.3 - 0.0111 preFlex + 0.172 preALF + 0.743 prePMax + 0.048 prePAve
- 0.0122 preLRMax - 0.09 Sex - 0.0949 Age

162 cases used, 21 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	28.32	10.10	2.81	0.006
preFlex	-0.01106	0.05066	-0.22	0.828
preALF	0.17242	0.05694	3.03	0.003
prePMax	0.7429	0.5616	1.32	0.188
prePAve	0.0482	0.5427	0.09	0.929
preLRMax	-0.01216	0.02220	-0.55	0.585
Sex	-0.092	1.598	-0.06	0.954
Age	-0.09494	0.08168	-1.16	0.247

S = 8.48589 R-Sq = 12.6% R-Sq(adj) = 8.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1602.68	228.95	3.18	0.004
Residual Error	154	11089.59	72.01		
Total	161	12692.28			

Stepwise regression model

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is 6wOx on 8 predictors, with N = 157
N (cases with missing observations) = 26 N (all cases) = 183

Step	1
Constant	9.958
preOx	0.549
T-Value	6.39
P-Value	0.000
S	7.98
R-Sq	20.86
R-Sq(adj)	**20.35**
Mallows Cp	0.3

Appendix E: Protocols (Chapter 5)

Local cell processing protocols (QMRI) adhered to:

H and E staining protocol

1. Stain in hematoxylin for 5 minutes
2. Rinse in running tap water for 20 minutes
3. Decolorize in acid alcohol for 1-3 seconds
4. Rinse in running tap water for 5 minutes
5. Immerse in Lithium Carbonate bath for 3 Seconds
6. Rinse in running tap water for 5 minutes
7. Counterstain in Eosin bath for 15 seconds
8. Dehydrate:
    a. Bathe in 90% ethanol for 3 minutes, air dry for 5 seconds
    b. Bathe in 95% ethanol for 3 minutes, air dry for 5 seconds
    c. Bathe in 100% ethanol (bath 1) for 3 minutes, air dry for 5 seconds
    d. Bathe in 100% ethanol (bath 2) for 3 minutes, air dry for 5 seconds
9. Clear in xylene (bath 1) for 5 minutes, air dry for 5 seconds
10. Clear in xylene (bath 2) for 5 minutes, air dry for 5 seconds

De-waxing protocol

1. Bathe in xylene (bath 1) for 5 minutes, then air dry for 5 seconds
2. Bathe in xylene (bath 2) for 2 minutes, then air dry for 5 seconds
3. Bathe in 100% ethanol (bath 1) for 2 minutes, then air dry for 5 seconds
4. Bathe in 100% ethanol (bath 2) for 2 minutes, then air dry for 5 seconds
5. Bathe in 90% ethanol for 2 minutes, then air dry for 5 seconds
6. Bathe in 80 % ethanol for 2 minutes, then air dry for 5 seconds
7. Bathe in 50% ethanol, for 2 minutes then air dry for 5 seconds
8. Rinse in distilled water for 5 minutes

Antigen retrieval protocol

1. Rinse for 1 minute in running tap water
2. Immerse in Sodium citrate and boil for 15 minutes (microwave)
3. Rinse in tepid running tap water for 15 minutes

Appendix F: Statistical output (Chapter 5)

Regression analysis of satellite cell PSI and post-operative change in power output

```
The regression equation is:

Improvement (W) 6/52-26/52 = - 2.33 + 2.89 cell PSI


16 cases used, 2 cases contain missing values

Predictor      Coef  SE Coef      T      P
Constant     -2.332    9.102  -0.26  0.802
Cell PSI      2.890    1.128   2.56  0.023

S = 11.9649   R-Sq = 31.9%   R-Sq(adj) = 27.0%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       1   939.2  939.2  6.56  0.023
Residual Error  14  2004.2  143.2
Total           15  2943.4
```

Regression analysis of satellite cell PSI and post-operative change in power-bodyweight ratio

```
The regression equation is:

Improvement (%BW) 6/52-26/52 = - 0.100 + 0.0459 cell PSI


16 cases used, 2 cases contain missing values

Predictor       Coef  SE Coef      T      P
Constant     -0.0996   0.1191  -0.84  0.417
Cell PSI     0.04589  0.01476   3.11  0.008

S = 0.156573   R-Sq = 40.8%   R-Sq(adj) = 36.6%


Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.23679  0.23679  9.66  0.008
Residual Error  14  0.34321  0.02451
Total           15  0.58000
```

Regression analysis of pre-operative power output and post-operative change in power output

```
The regression equation is:

Improvement (W) 6/52-26/52 = 15.5 + 0.0878 pre-op power


16 cases used, 2 cases contain missing values

Predictor           Coef  SE Coef     T      P
Constant          15.456    4.838  3.19  0.006
pre-op power     0.08781  0.07059  1.24  0.234

S = 13.7595   R-Sq = 10.0%   R-Sq(adj) = 3.5%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       1   292.9  292.9  1.55  0.234
Residual Error  14  2650.5  189.3
Total           15  2943.4
```

Regression analysis of pre-operative power output and post-operative change in power-bodyweight ratio

```
The regression equation is:

Improvement (%BW) 6/52-26/52 = 0.214 + 0.0677 pre-op (%BW)


16 cases used, 2 cases contain missing values

Predictor          Coef  SE Coef     T      P
Constant        0.21363  0.06964  3.07  0.008
St to BW ratio  0.06766  0.09037  0.75  0.466

S = 0.199584   R-Sq = 3.8%   R-Sq(adj) = 0.0%


Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.02233  0.02233  0.56  0.466
Residual Error  14  0.55767  0.03983
Total           15  0.58000
```

Stepwise regression model of change in power output

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is Improvement (W) 6/52-6/12 on 2 predictors, with N = 16
N (cases with missing observations) = 2 N (all cases) = 18

Step	1
Constant	-2.332
Cell PSI	2.9
T-Value	2.56
P-Value	0.023
S	12.0
R-Sq	31.91
R-Sq(adj)	**27.04**
Mallows Cp	2.2

Stepwise regression model of change in power-bodyweight ratio

Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05

Response is Improvement (%BW) 6/52-6/12 on 2 predictors, with N = 16
N (cases with missing observations) = 2 N (all cases) = 18

Step	1
Constant	-0.09963
Cell PSI	0.046
T-Value	3.11
P-Value	0.008
S	0.157
R-Sq	40.83
R-Sq(adj)	**36.60**
Mallows Cp	1.0

Appendix G: Protocols (Chapter 6)

RNA purification protocol

Total RNA Isolation **User Manual**

MACHEREY-NAGEL

NucleoSpin® RNA II / NucleoSpin® RNA L

July 2010 / Rev. 12

Total RNA purification from cultured cells and tissue with NucleoSpin® RNA II

Before starting the preparation:

Check if Wash Buffer RA3 and rDNase were prepared according to section 3.

1 Homogenize sample

Disrupt up to **30 mg** of **tissue** (for sample amounts see section 2.2; for homogenization methods see section 2.3). Up to **5 x 10_6** eukaryotic **cultured cells** can be collected by centrifugation and lysed by addition of Buffer RA1 directly.

2 Lyse cells

Add **350 µL Buffer RA1** and **3.5 µL ß-mercaptoethanol** (ß-ME) to the cell pellet or to ground tissue and vortex vigorously.

For appropriate sample and lysis buffer amounts see section 2.2.

3 Filtrate lysate

Reduce viscosity and clear the lysate by filtration through **NucleoSpin® Filter (violet ring)**: Place NucleoSpin® Filter in a Collection Tube (2 mL), apply the mixture, and centrifuge for **1 min** at **11,000 x *g***.

The lysate may be passed alternatively ε 5 times through a 0.9 mm needle (20 gauge) fitted to a syringe. In case of visible pellet formation (depending on sample amount and nature) transfer supernatant without any formed pellet to a new 1.5 mL microcentrifuge tube (not supplied). Important:

To process higher amounts of cells (> 1 x 10_6) or tissue (> 10 mg), the lysate should first be homogenized using the 0.9 mm needle (20 gauge), followed by filtration through NucleoSpin® Filters.

4 Adjust RNA binding conditions

Discard the NucleoSpin® Filter and add **350 µL ethanol (70 %)** to the homogenized lysate and mix by pipetting up and down (5 times).

Alternatively, transfer flow-through into a new 1.5 mL microcentrifuge tube (not provided), add **350 µL ethanol (70%)**, and mix by vortexing (2 x 5 s).

After addition of ethanol a stringy precipitate may become visible which will not affect the RNA isolation. Be sure to disaggregate any precipitate by mixing and load all of the precipitate on the column as described in step 5. Do not centrifuge the ethanolic lysate before loading it onto the column in order to avoid pelleting the precipitate.

5 Bind RNA

For each preparation take one **NucleoSpin® RNA II Column (light blue ring)** placed in a Collection Tube. Pipette lysate up and down 2 – 3 times and **load the lysate** to the column. Centrifuge for **30 s** at **11,000 x *g***. Place the column in a new Collection Tube (2 mL).

Maximal loading capacity of NucleoSpin® RNA II Columns is 750 µL. Repeat the procedure if larger volumes are to be processed.

6 Desalt silica membrane

Add **350 µL MDB** (Membrane Desalting Buffer) and centrifuge at **11,000 x *g*** for **1 min** to dry the membrane.

Salt removal will make the following rDNase digest much more effective. If the column outlet has come into contact with the flow-through for any reason, discard the flow-through and centrifuge again for 30 s at 11,000 x g.

7 Digest DNA

Prepare DNase reaction mixture in a sterile 1.5 mL microcentrifuge tube (not provided): For each isolation, add **10 µL reconstituted rDNase** (also see section 3) to **90 µL Reaction Buffer for rDNase**. Mix by flicking the tube. Apply **95 µL DNase reaction mixture** directly onto the center of the silica membrane of the column. Incubate at **room temperature** for **15 min**.

8 Wash and dry silica membrane

1st wash: Add **200 µL Buffer RA2** to the NucleoSpin® RNA II Column. Centrifuge for **30 s** at **11,000 x *g***. Place the column into a new Collection Tube (2 mL). *Buffer RA2 will inactivate the rDNase.*

2nd wash: Add **600 µL Buffer RA3** to the NucleoSpin® RNA II Column. Centrifuge for **30 s** at **11,000 x *g***. Discard flowthrough and place the column back into the Collection Tube.

Note: Make sure that residual buffer from the previous wash step is washed away with Buffer RA3.

3rd wash: Add **250 µL Buffer RA3** to the NucleoSpin® RNA II Column. Centrifuge for **2 min** at **11,000 x *g*** to dry the membrane completely. Place the column into a nucleasefree Collection Tube (1.5 mL, supplied).

If for any reason, the liquid level in the Collection Tube has reached the NucleoSpin® RNA II Column after centrifugation, discard flow-through, and centrifuge again. Note: Make sure that residual buffer from the previous wash step is washed away with Buffer RA3.

9 Elute RNA

Elute the RNA in **60 µL RNase-free H_2O**, (supplied) and centrifuge at **11,000 x *g***. for **1 min**.

If higher RNA concentrations are desired, elution can be done with 40 µL. Overall yield, however, will decrease when using smaller volumes. For further alternative elution procedures see section 2.4.

High Capacity cDNA Reverse Transcription Kits
For 200 and 1000 Reactions

Applied Biosystems

Protocol

Using the High Capacity cDNA Reverse Transcription Kits

RNA Template Guidelines

For optimal performance of the High Capacity cDNA Reverse Transcription Kits, Applied Biosystems recommends using RNA that is:

• Free of inhibitors of reverse transcription and PCR
• Dissolved in PCR-compatible buffer or water
• Free of RNase activity

Note: If you suspect that the RNA contains RNase activity, add RNase Inhibitor to the reverse transcription reaction at a final concentration of 1.0 U/µL.

Input Amount of Total RNA

Use up to 2 µg of total RNA per 20-µL reaction.

Preparing the 2× Reverse Transcription Master Mix

Prepare the 2× RT master mix using the kit components before preparing the reaction plate.

To prepare the 2× RT master mix (per 20-µL reaction):
1. Allow the kit components to thaw on ice.
2. Referring to the table below, calculate the volume of components needed to prepare the required number of reactions.

Note: Prepare the RT master mix on ice.

Component	Volume/Reaction (µL)	
	Kit with RNase Inhibitor	Kit without RNase Inhibitor
10× RT Buffer	2.0	2.0
25× dNTP Mix (100 mM)	0.8	0.8
10× RT Random Primers	2.0	2.0
MultiScribe™ Reverse Transcriptase	1.0	1.0
RNase Inhibitor	1.0	—
Nuclease-free H_2O	3.2	4.2
Total per Reaction	10.0	10.0

IMPORTANT! Include additional reactions in the calculations to provide excess volume for the loss that occurs during reagent transfers.

3. Place the 2× RT master mix on ice and mix gently.

Preparing the cDNA Reverse Transcription Reactions

To prepare the cDNA RT reactions:

1. Pipette 10 µL of 2× RT master mix into each well of a 96-well reaction plate or individual tube.
2. Pipette 10 µL of RNA sample into each well, pipetting up and down two times to mix.
3. Seal the plates or tubes.
4. Briefly centrifuge the plate or tubes to spin down the contents and to eliminate any air bubbles.
5. Place the plate or tubes on ice until you are ready to load the thermal cycler.

Performing Reverse Transcription

To perform reverse transcription:

1. Program the thermal cycler conditions using one of the thermal cyclers listed in Table 3 on page 4.

IMPORTANT! These conditions are optimized for use with the High Capacity cDNA Reverse Transcription Kits.

	Step 1	Step 2	Step 3	Step 4
Temperature (°C)	25	37	85	4
Time	10 min	120 min	5 min	□

2. Set the reaction volume to **20 µL**.
3. Load the reactions into the thermal cycler.
4. Start the reverse transcription run.

Storing cDNA Reverse Transcription Reactions

You can store cDNA RT plates or tubes prepared using the High Capacity cDNA Reverse Transcription Kits for short-term or long-term storage.

Storage Duration	Storage Temperature (°C)
Short-term (up to 24 hours before use)‡	2 to 6
Long-term	–15 to –25

‡ For prolonged storage at 2 to 6 °C, add EDTA to a final concentration of 1 mM to chelate cations and to prevent nucleic acid degradation.

IMPORTANT! If required, briefly centrifuge the archive plates or tubes before storing to spin down the contents and to eliminate any air bubbles.

Appendix H: Statistical output (Chapter 6)

Regression of Pax7 expression and post-operative change in power output

```
The regression equation is:

Improvement (W) = 11.0 + 33.0 Pax-7


11 cases used, 1 cases contain missing values

Predictor    Coef  SE Coef     T      P
Constant   10.965    9.071  1.21  0.258
Pax-7       33.02    15.54  2.13  0.062

S = 19.6830   R-Sq = 33.4%   R-Sq(adj) = 26.0%


Analysis of Variance

Source          DF      SS      MS     F      P
Regression       1  1749.9  1749.9  4.52  0.062
Residual Error   9  3486.8   387.4
Total           10  5236.7
```

Regression of Pax7 expression and post-operative change in power-bodyweight ratio

```
The regression equation is:

Improvement (%BW) = 0.0842 + 0.345 Pax-7


Predictor     Coef  SE Coef     T      P
Constant   0.08419  0.05205  1.62  0.140
Pax-7      0.34466  0.08916  3.87  0.004

S = 0.112950   R-Sq = 62.4%   R-Sq(adj) = 58.2%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       1  0.19064  0.19064  14.94  0.004
Residual Error   9  0.11482  0.01276
Total           10  0.30545
```

Regression of NCAM expression and post-operative change in power output

```
The regression equation is:

Improvement (W) = 3.80 + 26.0 NCAM


11 cases used, 1 cases contain missing values

Predictor    Coef  SE Coef     T      P
Constant    3.797    6.450  0.59  0.571
NCAM       25.963    5.932  4.38  0.002

S = 13.6377   R-Sq = 68.0%   R-Sq(adj) = 64.5%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  3562.9  3562.9  19.16  0.002
Residual Error   9  1673.9   186.0
Total           10  5236.7
```

Regression of NCAM expression and post-operative change in power-bodyweight ratio

```
The regression equation is:

Improvement (%BW) = 0.0678 + 0.201 NCAM


11 cases used, 1 cases contain missing values

Predictor     Coef  SE Coef     T      P
Constant   0.06779  0.04766  1.42  0.189
NCAM       0.20123  0.04384  4.59  0.001

S = 0.100781   R-Sq = 70.1%   R-Sq(adj) = 66.7%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       1  0.21404  0.21404  21.07  0.001
Residual Error   9  0.09141  0.01016
Total           10  0.30545
```

Regression of CD34 expression and post-operative change in power output

```
The regression equation is:

Improvement (W) = 28.2 - 8.9 Cd34


11 cases used, 1 cases contain missing values

Predictor   Coef  SE Coef      T      P
Constant   28.24    14.60   1.93  0.085
CD34       -8.87    41.72  -0.21  0.836

S = 24.0614   R-Sq = 0.5%   R-Sq(adj) = 0.0%


Analysis of Variance

Source          DF      SS     MS     F      P
Regression       1    26.2   26.2  0.05  0.836
Residual Error   9  5210.6  579.0
Total           10  5236.7
```

Regression of CD34 expression and post-operative change in power-bodyweight ratio

```
The regression equation is:

Improvement (%BW) = 0.243 - 0.021 Cd34


11 cases used, 1 cases contain missing values

Predictor     Coef  SE Coef      T      P
Constant    0.2428   0.1118   2.17  0.058
Cd34       -0.0211   0.3194  -0.07  0.949

S = 0.184182   R-Sq = 0.0%   R-Sq(adj) = 0.0%


Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.00015  0.00015  0.00  0.949
Residual Error   9  0.30531  0.03392
Total           10  0.30545
```

Stepwise regression model of change in power output

```
Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05


Response is Improvement (W) on 3 predictors, with N = 1O
N (cases with missing observations) = 1 N (all cases) = 11


Step              1
Constant      3.797

NCAM           26.0
T-Value        4.38
P-Value       0.002

S              13.6
R-Sq          68.04
R-Sq(adj)     64.48
```

Stepwise regression model of change in power-bodyweight ratio

```
Alpha-to-Enter: 0.05, Alpha-to-Remove: 0.05


Response is Improvement (%BW) on 3 predictors, with N = 10
N (cases with missing observations) = 1 N (all cases) = 11


Step               1
Constant      0.06779

NCAM            0.201
T-Value          4.59
P-Value         0.001

S               0.101
R-Sq            70.07
R-Sq(adj)       66.75
Mallows Cp        2.2
```

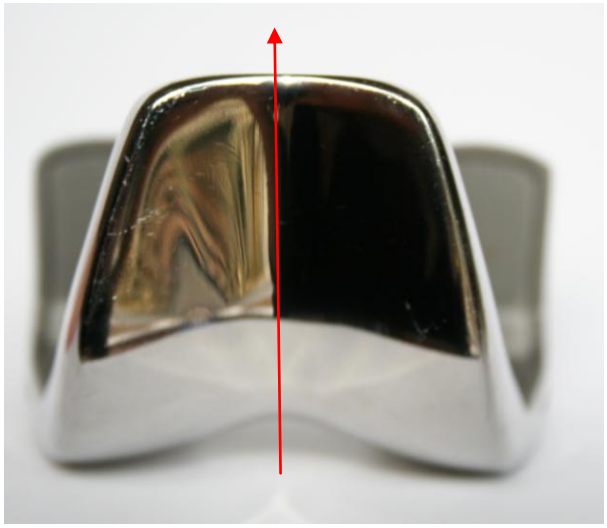
Appendix I: Implant design differences

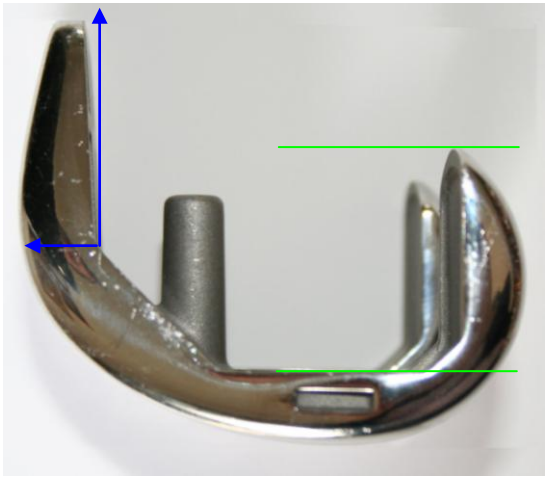
Highlighted design differences between the implants used.

Further design differences employed in the Triathlon implant compared to the Kinemax implant can be visualised in the following images:

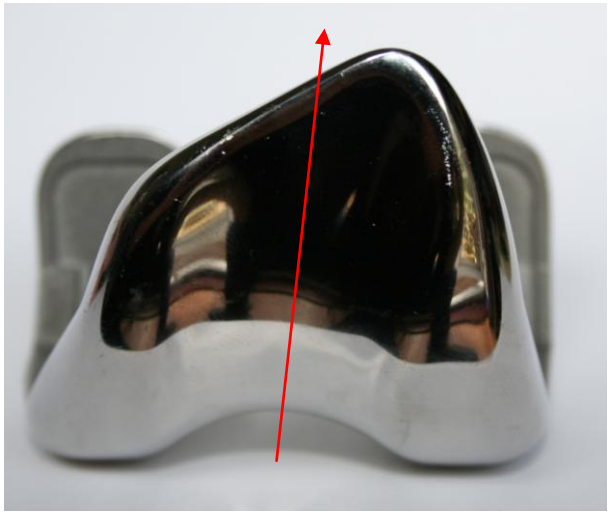
1 - Sided patello-femoral grove: AP view (annotated with red arrows) and axial view.
2 - Reduced anterior bulk of implant, but increased angle of anterior flange: lateral view (annotated with blue arrows) and axial view.
3 - Shorter posterior condylar offset: lateral view (annotated with green bars)

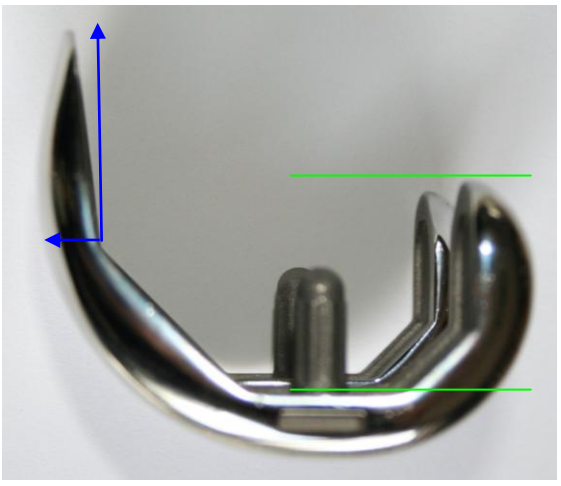
Kinemax Total Knee Prosthesis





Triathlon Total Knee Prosthesis





Images of equivalent sized prostheses. Individual anteroposterior and lateral views. Photographs taken with a Canon DSLR camera.

Images of equivalent sized prostheses. Combined axial view (Kinemax prosthesis to the left, Triathlon prosthesis to the right). Photographs taken with a Canon DSLR camera.

Bibliography

Albertz, B., Bery, D., Lewis, J., Raff, M., Roberts, K., Watson. J., 1994, *Molecular Biology of The Cell*, 4th Ed, Garland Publishing Inc, New York.

American Academy of Orthopaedic Surgeons, 2010, Knee replacement implants. Available at: http://www.orthoinfo.org/topic.cfm?topic=A00221 [Accessed 24 March 2011]

Ashby, E., Grocott M.P.W., Haddad, F.S., 2008. Outcome measures for orthopaedic interventions on the hip. *J Bone Joint Surg [Br]*, 90-B, pp.545-9.

Asmussen, E., Heeboll-Nielson, K., 1962. Isometric muscle strength in relation to age in men and women. *Ergonomics*, 5, pp.167-9.

Aspden, R.M., 2008. Osteoarthritis: a problem of growth not decay? *Rheumatology (Oxford)*, 47(10), pp.1452-60.

Barton, E., and Morris, C., 2003. Mechanisms and strategies to counter muscle atrophy. *J Gerontol Med Sci*, 58A, pp.923-6.

Back, D.L., Cannon, S.R., Hilton, A., Bankes, M.J., Briggs, T.W., 2001, The Kinemax total knee arthroplasty. Nine years' experience. *J Bone Joint Surg [Br]*, 83-B(3), pp.359-63.

Bassey, E.J., Fiatarone, M.A., O'Neill, E.F., Evans, W.J., Lipsitz, L.A. 1992. Leg extensor power and functional performance in very old men and women. *Clinical Science*, 82, pp.321-7.

Bassey, E.J., and Short, A.H., 1990. A new method for measuring power output in a single leg extension: feasibility, reliability and validity. *Eur J Appl Physiol*, 60, pp.385-90.

Barker, K.L., Lamb, S.E., Simpson, A.H.R.W., 2004. Functional recovery in patients with non-union treated with the Ilizarov technique. *J Bone Joint Surg [Br]*, 86-B, pp.81-5.

Beals, C., Lampman, R., Banwell, B., Brannstein, E., Albers, J., Castor, C.W., 1985. Measurement of exercise tolerance in patients with rheumatoid arthritis and osteoarthritis. *J Rheumatol*, 12, pp.458-61.

Bean, J.F., Leveille, S.G., Kiely, D.K., Bandinelli, S., Guralnik, J.M., Ferrucci, L., 2003. A comparison of leg power and leg strength within the InCHIANTI study: Which influences mobility more? *J Gerontol A Boil Sci Med Sci*, 58, pp.728-33.

Bear, W.S., 1918. Arthroplasty with the aid of an animal membrane. *Am J Orthop Surg*, 16, pp.171-199.

Beard, D.J., Knezevic, K., Al-Ali, S., Dawson, J., Price, A.J., 2010. The use of outcome measures relating to the knee *Othropaedics & Trauma*, 24(4), pp.309-16.

Beauchamp, J,R,, Heslop, L., Yu, D.S., Tajbakhsh, S., Kelly, R.G., Wernig, A., Buckingham, M.E., Partridge, T.A., Zammit, P.S., 2000. Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. *J Cell Biol,* 151(6), pp.1221-34.

Bellamy, N., 1993. *Musculoskeletal Clinical Metrology*. London. Kluwer Academic Publishing.

Bellamy, N., Kirwan, J., Boers, M., Brooks, P., Strand, V., Tugwell, P., Altman, R., Brandt, K., Dougados, M., Lequesne, M., 1997. Recommendations for a core set of outcome measures for future phase III clinical trials in hip knee and hand osteoarthritis. Consensus development at OMERCAT III. *J Rheumatol*, 24, pp.799-802

Berman, A.T., Bosacco, S.J., Isralite, C., 1991. Evaluation of total knee arthroplasty using isokinetic testing. *Clin Orthop Relat Res*, 271, pp.106-13.

Bischoff, R., 1975. Regeneration of single skeletal muscle fibers in vitro. *Anat Rec*, 182, pp.215–35.

Black, N., and Jenkinson, C., 2009. How can patients' views of their care enhance quality improvement? BMJ, 339:b2495.
doi: 10.1136/bmj.b2495 (Published 2 July 2009).

Boldrin, L., Muntoni, F., and Morgan, J., 2010. Are Human and Mouse Satellite Cells Really the Same? *J. Histochem. Cytochem, ;*58(11), pp.941–55.

Boonstraa, M.C., De Waal Malefijtb, M.C., Verdonschota, N., 2008. How to quantify knee function after total knee arthroplasty? *Knee*, 15(5), pp.390-5.

Bourne, R.B., McCalden, R.W., MacDonald, S.J., Mokete, L., Guerin, J., 2007. Influence of patient factors on TKA outcomes at 5 to 11 years followup. *Clin Orthop Relat Res*, 464, pp.27-31.

Bream, E., Charman, S.C., Clift, B., Murray, D., Black, N., 2010. Relationship between patients' and clinicians' assessments of health status before and after knee arthroplasty. *Qual Saf Health Care*, 19(6):e6. Epub. Jul 1.

Bream, E., and Black. N., 2009, What is the relationship between bateints' and clinicains' reports of the outcomes of surgery? *J Health Serv Res Policy,* 14(3), pp.174-82.

British Orthopaedic Association / British Association for Surgery of the Knee position statement, 2010, *Knee Replacement: A Guide to Good Practice*. Available at: http://www.library.nhs.uk/integratedSearch/viewResource.aspx?resID=113744 [Accessed 24 March 2011]

Brooks, S.V., and Faulkner, J.A., 1994. Skeletal muscle mass weakness in old age: underlying mechanisms. *Med Sci Sports Exerc*, 26, pp.432-9

Brosseau, L., Balmer, S., Tousignant, M., O'Sullivan, J.P., Goudreault, C., Goudreault, M., Gringras, S., 2001. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. *Arch Phys Med Rehabil*, 82(3), pp.396-402.

Brosseau, L., Tousignant, M., Budd, J., Chartier, N., Duciaume, L., Plamondon, S., O'Sullivan, J.P., O'Donoghue, S., Balmer, S., 1997. Intratester and intratester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. *Physiother Res Int*, 2(3), pp. 150-66.

Brown, J.E., McGaw, W.H., Shaw, D.T., 1958. Use of cutis as an interposing memberane in arthroplasty of the knee. Report of cases. *J Orthop Surg* 40-A, pp.103.

Buckingham, M., 2007. Skeletal muscle progenitor cells and the role of the Pax genes. *C.R. Biologies*, 330, pp.530–3.

Campbell, N. 1996, *Biology*, 4th Ed, California: Benjamin Cummings Publishing

Carlson, B.M., and Faulkner, J.A., 1989. Muscle transplantation between young and old rats: Age of host determines recovery. *Am. J. Physiol*. 256, C1262–C1266.

Cerletti, M., Jurga, S., Witczak, C.A., Hirshman, M.F., Shadrach, J.L., Goodyear, L.J., Wagers, A.J., 2008. Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell*, 134, pp.37–47.

Charifi, N., Kadi, F., Feasson, L., Denis, C., 2003. Effects of endurance training on satellite cell frequency in skeletal muscle of old men. *Muscle Nerve*, 28-1, pp.87-92.

Churchill, D., Incavo, S., Johnson, C., Beynnon, B.D. 1998. The transepicondylar axis approximates the optimal flexion axis of the knee. *Clin Orthop*, pp.111-8.

Collins, C.A., Olsen, I., Zammit, P.S., Heslop, L., Petrie, A., Partridge, T.A., Morgan, J.E., 2005. Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell*, 122-2, pp.289-301.

Conboy, M., Conboy, M.J., Wagers, A.J., Girma, E.R., Weissman, I.L., Rando, T.A., 2005. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature*, 433, pp.760–764

Crameri, R.M., Langberg, H., Magnusson, P., Jensen, C.H., Schroder, H.D., Olesen, J.L., Suetta, C., et al., 2004. Changes in satellite cells in human skeletal muscle after a single bout of high intensity exercise. *J Physiol* 558, pp.333–340

Culliford, D.J., Maskell, J., Beard, D.J., Murray, C.W., Price, A.J., Arden, N.K., 2010. Temporal trends in Hip and knee replacement in the United Kingdom 1991 to 2006, *J Bone Joint Surg [Br]*, 92-B pp.130-5.

Davies, A.P., 2002. Rating systems for total knee replacement. *Knee*, 9, pp.261-6.

Dawson, J., Fitzpatrick, R., Murray, D., Carr, A., 1998. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg [Br]*, 78-B, pp.185-90.

Decary, S., Hamida, C.B., Mouly, V., Barbet, J.P., Hentati, F., Butler-Browne, G.S., 2000 Shorter telomeres in dystrophic muscle consistent with extensive regeneration in young children. *Neuromuscul Disord*, 10-2, pp.113-20.

Decary, S., Mouly, V., Hamida, C.B., Sautet, A., Barbet, J.P., Butler-Browne GS., 1997. Replicative potential and telomere length in human skeletal muscle: implications for satellite cell-mediated gene therapy. *Hum Gene Ther*, 8-12, pp.1429-38.

Department of Health. 2008. High quality health care for all. NHS next stage review. London: DoH.

Dennis, D.A., Komisktek, R.D., Scuderi, G.R., Zingde, S., 2007. Factors affecting flexion after total knee arthroplasty. *Clin Orthop Relat Res*, 464, pp.53-60.

D'Lima, D., Poole, C., Chadha, H., Hermida, J.C., Mahar, A., Colwell, C.W., 2001. Quadriceps moment arm and quadriceps forces after total knee arthroplasty. *Clin Orthop Relat Res*, 392, pp.213-20.

Dunbar, M.J., Robertsson, O., Ryd, L., Lidgren, L., 2001. Appropriate questionnaires for knee arthroplasty: Results of a survey of 3600 patients from the sweedish knee arthroplasty registry. *J Bone Joint Surg [Br]*, 83-B, pp.339-44.

Dutta, C., and Hadley, E., 1995. The significance of sarcopenia in old age. *J Gerontol A Biol Sci Med Sci*, 50, pp.1-4.

Ethgen, O., Bruyere, O., Richy, F., Dardennes, C., Reginster, J.Y., 2004. Health related quality of life in total hip and total knee arthroplasty: A qualitative and systematic review of the literature. *J Bone Joint Surg [Am]*, 86-A, pp.963–74.

Evans, W.J., 2000. Exercise strategies should be designed to increase muscle power [editorial]. *J Gerontol A Biol Sci Med Sci*, 55, pp.M309-10.

Ferreira, R., Neuparth, M.J., Ascensao, A., Magalhaes, J., Vitorino, R., Duarte, J.A., Amado, F., 2006. Skeletal muscle atrophy increases cell proliferation in mice gastrocnemius during the first week of hindlimb suspension. *Eur J Appl Physiol* 97, pp340–346

Fayers, P.M., Machin, D., 2000. Quality of life: assessment, analysis and interpretation. Chichester: John Wiley & Sons Ltd

Ferguson, M., 1861. Excision of the knee joint. Recovery with a false joint and useful limb. *Med Times Gaz*, 1, pp.601.

Franklin, P.D., Li, W., Ayres, D.C., 2008. The Chitranjan Ranawat Award: Functional Outcome after Total Knee Replacement Varies with Patient Attributes. *Clin Orthop Relat Res*, 466, pp.2597-604.

Gapeyeva, H., Buht, N., Peterson, K., Ereline, J., Haviko, T., Paasuke, M., 2007. Quadriceps femoris muscle voluntary isometric force production and relaxation characteristics before and 6 months after unilateral total knee arthroplasty in women. *Knee Surg Sports Traumatol Arthrosc*. 15, pp.202-11.

Garrat, A.M., Brealey, S., Gillespie, W.J., 2004. Patient assessed health instruments for the knee: a structured review. *Rheumatology (Oxford)*, 43, pp.1414-23.

Gibbs, A.N., Green, G.A., Taylor, J.G., 1979. A comparison of the Freeman-Swanson (ICLH) and Wallidus prosthesis in total knee replacement. *J Bone Joint Surg [Br]*, 61-B, pp.358.

Gluck T. (1894) Autoplastik-Transplantation-Implantation von Fremdkorpern. *Berl klm Wschr*, 27, pp.421-7.

Gomez-Barrena, E., Fernandez-Garcia, C., Fernandez-Bravo, A., Cutillas-Ruiz, R., Bermejo-Fernandez, G., 2010. Functional performance with a single-radius femoral design total knee arthroplasty. *Clin Orthop Relat Res*, 468, pp.1214-20.

Greene, K.A., and Schurman, J.R., 2008. Quadriceps muscle function in primary total knee arthroplasty. *J Arthroplasty* , 23(7) Suppl. 1.

Grounds, M.D., 1998. Age-associated changes in the response of skeletal muscle to exercise and regeneration. *Ann N.Y. Acad Sci*, 854, pp.78-91.

Gunston, F.H., 1971. Polycentric knee arthroplasty. *Clin Orthop*, 94, pp.128.

Haber, J.E., Thorburn, P.C., 1984. Healing of broken dicentric chromosomes in yeast. *Genetics*, 106, pp.207-26

Hall, J., Banks, G., Chamberlain, J., Olwin, B., 2010. Prevention of Muscle Aging by Myofiber-Associated Satellite Cell Transplantation. *Sci Transl Med*, 2, pp.57.

Hall, J., Copp, S., Adelson, W., D'Lima, D., Colwell, C.W., 2008. Extensor mechanism function in single-radius vs multiraduis femoral components for total knee arthroplasty. J arthroplasty, 23(2), pp.216-9.

Harley, C.B., Futcher, A.B., Greider, C.W., 1990 Telomeres shorten during ageing of human fibroblasts. *Nature*, 345-6274, pp.458-60.

Hawke, T.J., Garry, D.J., 2001. Myogenic satellite cells: physiology to molecular biology. *J Appl Physiol*, 91-2, pp.534-51.

Hawker, G.A., 2006. Who, when, and why total joint replacement surgery? The patient's perspective. *Curr Opin Rheumatol*, 18, pp.526–30.

Hikida, R.S., Walsh, S., Barlyski, N., Campos, G., Hagerman, F.C., Staron, R.S., 1998. Is hypertrophy limited in elderly muscle fibres? A comparison of elderly and young strength-trained men. *Basic Appl Myol*, 8, pp.419-27.

Hollister A, Jatana S, Sing A. Sullivan WW, Lupichuk AG. (1993) The axes of rotation of the knee. Clin Orthop;290:259-68.

Hoshino, A., 1997. Influence of design on the range of motion after total knee arthroplasty. In: Niwa, S., Yoshino, S., Kurosaka, M., Eds. *Reconstruction of the knee joint*. Hong Kong: Springer-Verlag Tokyo; p 324

Huang, C.H., Liau, J.J., Cheng, C.K., 2007. Fixed or mobile-bearing total knee arthroplasty. *J Orthop Surg Res*, 2, pp.1.

Huang, C., Cheng, C., Lee, Y., Lees, K., 1996. Muscle strength after successful total knee replacement: a 6 to 13 year follow-up. *Clin Orthop*, 328, pp.147-54.

Hurley, M., 1999. The role of muscle weakness in the pathogenesis of osteoarthritis *Rhuem Dis Clin North Am*, 25, pp.283-98.

Hurley, M., and Scott, D., 1998. Improvement is quadriceps sensorimotor function and disability of patients with knee osteoarthritis following a clinically practicable exercise regime. *Br J Rheumatol*, 25, pp.2432-9

Illa, I., Leon-Monzon, M., Dalakas, M.C., 1992. Regenerating and denervated human muscle fibers and satellite cells express neural cell adhesion molecule recognized by monoclonal antibodies to natural killer cells. *Ann Neurol*, 31, pp.46–52

Insall, J., Ranawat, C.S., Scott, W.N., Walker, P., 1976. Total condylar knee replacment: preliminary report. *Clin Orthop Relat Res*, 120, pp.149-54.

Insall, J.N., Dorr, L.D., Scott, R.D., Scott, W.N., 1989. Rationale of the Knee Society clinical rating system. *Clin Orthop*, 248, pp.13-4.

Jacobs, W., Clement, D.J., Wymenga, A.A.B., 2007a. Retention versus sacrifice of the posterior cruciate ligament in total knee replacement for treatment of osteoarthritis and rheumatoid arthritis. *Cochrane Database of Systematic Reviews*, Issue 4.

Jacobs, W., Anderson, P.G., van Limbeek, J., Wymenga, A.A.B., 2007b. Mobile bearing vs. fixed bearing prostheses for total knee arthroplasty for post-operative functional status in patients with osteoarthritis and rheumatoid arthritis. *Cochrane Database of Systematic Reviews* 2001, Issue 4.

Jakobsen, T.L., Christensen, M., Christensen, .SS., Olsen, M., Bandholm, T., 2010. Reliability of knee joint range of motion and circumference measurements after total knee arthroplasty: does tester experience matter? *Physiother Res Int*, 15(3), pp.126-34.

Jensen, M.P., Karolyn, P., 2001. Self-report scales and procedures for assessing pain in adults. In: Turk, D.C., Melzack, R., Eds. *Handbook of Pain Assessment*. 2nd Ed. New York, NY: Guilford Press.

Jensen, M., Turner, L., Turner, J., Romano, J.M., 1996. Use of multiple-item scale for pain intensity measurement in chronic pain patients. *Pain*, 67, pp.35-40.

Jones, D.A., Round, J.M., 1990. *Skeletal muscle in health and disease: A textbook of muscle physiology* Manchester: Manchester University Press.

Jones, G.B., 1969, Arthroplasty of the knee by the Wallidus prosthesis. *J Bone Joint Surg [Br]*, 36-B, pp.553-60.

Juni, P., Reichenbach, S., Dieppe, P., 2006. Osteoarthritis: rational approach to treating the individual. *Best Pract Res Clin Rheumatol*, 20, pp.721–40.

Kadi, F., and Posnot, E., 2010. The biology of satellite cells and telomeres in human skeletal muscle: effects of aging and physical activity. *Scand J Med Sci Sports*, 20, pp.39-48.

Kadi, F., Charifi, N., Henriksson, J., 2006. The number of satellite cells in slow and fast fibres from human vastus lateralis muscle. *Histochem Cell Biol*, 126, pp.83–7.

Kadi, F., Charifi, N., Denis, C., Lexell, J., Andersen, J.L., Schjerling, P., Olsen, S., et al. 2005. The behaviour of satellite cells in response to exercise: what have we learned from human studies? *Pflugers Arch* 451, pp.319–327

Kadi, F., Charifi, N., Denis, C., Lexell, J., 2004, Satellite cells and myonuclei in young and elderly women and men. *Muscle Nerve*, 29-1, pp.120-7.

Kadi, F., Thornell, L.E., 2000. Concomitant increases in myonuclear and satellite cell content in female trapezius muscle following strength training. *Histochem Cell Biol*, 113, pp.99–103

Kadi, F., Eriksson, A., Holmner, S., Thornell, L.E., 1999. Effects of anabolic steroids on the muscle cells of strength-trained athletes. *Med Sci Sports Exerc*, 31-11, pp.1528-34.

Katz, B., 1961. The terminations of the afferent nerve fibre in the muscle spindle of the frog. Philos. *Trans. R. Soc. Lond*. 243, pp.221-40.

Kauhanen, S., Salmi, A., von Boguslawski., 2003. Satellite cell proliferation, reinnervation and revascularisation in human free microvascular muscle flaps. *J Surg Res*, 115, pp.191-9.

Kurtz, S., Ong, K., Lau, E., Mowat, F., Halpern, M., 2007. Projections of Primary and Revision Hip and Knee Arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg [Am]*, 89-A, pp.780–5.

Laguens, R., 1963. Satellite cells of skeletal muscle fibers in human progressive muscular dystrophy. *Virchows Arch Pathol Anat Physiol Klin Med*, 336, pp.564–9

Lamb, S.E., Frost, H., 2003. Recovery of mobility after knee arthroplasty. *J Arthroplasty*, 18, pp.575-82.

Larson, L., 1978. Morphological and functional characteristics of the aging skeletal muscle in man. A cross-sectional study. *Acta Physiol Scand*, 457, pp.1-36.

Lindstrom, M., Thornell, L.E., 2009. New multiple labelling method for improved satellite cell identification in human muscle: application to a cohort of power-lifters and sedentary men. *Histochem Cell Biol*, 132, pp.141–57

Lingard, E.A., Katz, J.N., Wright, E.A., Sledge, C.B., The Kinemax Outcomes Group, 2004. Predicting the outcome of total knee arthroplasty. *J Bone Joint Surg [Am]*, 86, pp.2179–86.

Lingard, E.A., Katz, J.N., Wright, R.J., Wright, E.A., Sledge, C.B., Kinemax Outcomes Group., 2001. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg [Am]*, 83-A(12), pp.1856-64.

Liow, R.Y.L., Murray, S.W., 1997. Which Primary Total Knee Replacement. A Review of Currently Available TKR in the United Kingdom. *Ann R Coll Surg Eng*, 79, pp.335-40.

Lexell, J., 1995. Human aging, muscle mass, and fibre type composition. *J Gerontol A Biol Sci Med Sci*, 50 Spec No:11-6.

Lexell, J., 1993. Ageing and human muscle: observations form Sweden. *Can J Appl Physiol*, 18, pp.2-18.

Lexell, J., Taylor, C.C., Sjöström, M., 1988 What is the cause of ageing atrophy? Total number, size and proportion of different fibre types studied in whole vastus lateralis muscle from 15 to 83 year old men. *J Neurol Sci*, 84(2-3), pp.275-94.

Lorentzen, J.S., Petersen, M.M., Brot, C., Madsen, O,R., 1999. Early changes in muscle strength after total knee arthroplasty. *Acta Orthop Scand*,70, pp.176-9.

Mackey, A.L., Esmarck, B., Kadi, F., Koskinen, S.O., Kongsgaard, M., Sylvestersen, A., Hansen, J.J., Larsen, G., Kjaer, M., 2007. Enhanced satellite cell proliferation with resistance training in elderly men and women. *Scand J Med Sci Sports*, 17-1, pp.34-42.

Mackintosh. D.L., 1966. Arthroplasty of the knee. *J Bone and Joint Surg [Br]* 48-B, pp.179.

Mahoney, O., McClung, C., Dela Rosa, M., Schmalzried, T.P., 2002. The effect of total knee arthroplasty design on extensor mechanism function. *J. Arthroplasty*, 17, pp.416.

Marquardt, O., 1950. Kinegelenksplastik mit VLA Stahplatte. Z Sch G. *Orthop*, 80, pp.140.

Mauro, A., 1961. Satellite cell of skeletal muscle fibers. *J Biophys Biochem Cytol*, 9, pp.493-5.

McCarthy, C.J., and Oldham, J.A., 2004. The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee. *Rheumato*l, 43, pp.514-7.

McEwan, H.M., Barnett, P.I., Bell, C.J., Farrar, R., Auger, D.D., Stone, M.H., Fisher, J., 2005. The influence of design, material and kinematics on the in vitro ear of total knee replacements. *J Biomech*, 38(2), pp.357-65.

Mechtersheimer, G., Staudter, M., Moller, P., 1992. Expression of the natural killer (NK) cell-associated antigen CD56 (Leu-19), which is identical to the 140-kDa isoform of N-CAM, in neural and skeletal muscle cells and tumors derived there from. *Ann N.Y. Acad Sci*, 650, pp.311–6.

Mizner, R.L., Petterson, S.C., Clements, K.E., Zeni, J.A., Irrgang, J.J., Snyder-Mackler, L., 2011. Measuring Functional Improvement After Total Knee Arthroplasty Requires Both Performance-Based and Patient-Report Assessments: A Longitudinal Analysis of Outcomes. *J Arthroplasty*, In Press, Corrected Proof, Available online 20 September 2010.

Mizner, R.L., Petterson, S.C., Stevens, J.E., Valdenborne, K., Snyder-Mackler, L., 2005a. Early Quadriceps Strength Loss After Total Knee Arthroplasty: The Contribution of Muscle Atrophy and Failure of Voluntary Muscle Activation. *J Bone Joint Surg [Am]*, 87, pp.1047-53.

Mizner, R.L., Petterson, S.C., Stevens, J.E., Axe, M.J., Snyder-Mackler, L., 2005b. Preoperative strength predicts functional ability one year after total knee arthroplasty *J Rheumatol*, 32, pp.1533-9.

Mizner, R.L., Snyder-Mackler L., 2005c. Altered loading during walking and sit-to-sand is affected by quadriceps weakness after total knee arthroplasty. *J Orthop Res*, 23, pp.1083-90.

Mont, M.A., Marker, D.R., Seyler, T.M., Jones, L.C., Kolisek, F.R., Hungerford, D.S., 2008. High-Impact Sports After Total Knee Arthroplasty. J Arthroplasty, 6 Suppl.1, pp.80-4.

Morgan, J.E., Partridge, T.A., 2003. Muscle satellite cells. *Int J Biochem Cell Biol*, 35-8, pp.1151-6.

Murray, D.W., Fitzpatrick, R., Roger,s K., Pandit, H., Beard, D.J., Carr, A.J., Dawson, J., 2007. The use of the oxford hip and knee scores. *J Bone Joint Surg [Br]*, 89-B, pp.1010-4.

Murray, D.W., Britton, A.R., Bulstrode, C.J.K., 1997. Loss to follow-up matters. *J Bone Joint Surg [Br]*, 79, pp.254-7.

Nailis, H., Coenye, T., Van Nieuwerburgh, F., Deforce, D., Nelis, H.J., 2006. Development and evaluation of different normalization strategies for gene expression studies in Candida albicans biofilms by real-time PCR. *BMC Mol Biol*, 7, pp.25.

National Collaborating Centre for Chronic Conditions. 2008. *Osteoarthritis: national clinical guideline for care and management in adults*. London: Royal College of Physicians.

National Joint Registry for England and Wales, 2010, *7th Annual Report*. Available at: www.njrcentre.org.uk. Accessed 24 March 2011.

Nielsen, J.S., McNagny, K.M., 2008. Novel functions of the CD34 family. *J Cell Sci*, 121, pp.3683-92.

NIH Consensus Statement on Total Knee Replacement. 2003. NIH Consens State Sci Statements. Dec 8-10; 20(1) 1-32

Noble, P.C., Conditt, M.A., Cook, K.F., Mathis, K.B., 2006. The John Insall Award: patient expectations affect satisfaction with total knee arthroplasty. *Clin Orthop Relat Res*, 452, pp.35-43.

Nolan, T., Hands, R.E., Bustin, S.A., 2006. "Quantification of mRNA using real-time RT-PCR.". *Nat. Protoc*, 1(3), pp.1559–82.

Nordesjo, L., Nordgren, B., Wigren, A., Kolstad, K., 1983. Isometric strength and endurance in patients with severe rheumatiod arthritis or osteoarthritis in the knee joints: A comparative study in healthy men and women. *Scand J Rheumatol*, 12, pp.152-6.

Nuesch, E., Dieppe, P., Reichenbach, S., Williams, S., Juni, P. 2011. All cause disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ, 342, pp.d1165

Nutton, R.W., van der Linden, M., Rowe, P.J., Gaston, P., Wade, F.A., 2008. A prospective randomised double-bind study of functional outcome and range of flexion following total knee replacement with the NexGen standard and high flexion components. *J Bone Joint Surg [Br]*, 90-B, pp.37-42.

O'Connor, J.J., Shercliff, T.L., Biden, E., Goodfellow, J.W., 1989. The geometry of the knee in the sagittal plane. *Proc Inst Mech Eng H*., 203(4), pp.223-33.

Pakos, E., Ntzani, E., Trikalinos, T., 2005. Patellar resurfacing in total knee arthroplasty. A meta-analysis. *J Bone Joint Surg [Am]*, 87, pp.1438-45.

Panjabi, M.M., Goel, V,K., Walte,r S.D., 1982. Errors in kinematic parameters of a planar joint: Guidelines for optimal experimental design. *J Biomech*, 15, pp.537.

Perrot, S., Rozenberg, S., Moyse, D., et al. 2010. Comparison of daily, weekly or monthly pain assessments in hip and knee osteoarthritis. A 29-day prospective study. Joint Bone Spine, doi:10.1016/j.jbspin2010.11.009.

Philip, B.K., 1990. Parametric statistics for evaluation of the visual analogue scale. *Anesth Analg*, 71, pp.710.

Pinskerova, V., Iwaki, H., Freeman, M.A., 2000. The shapes and relative moments of the femur and tibia at the knee. *Orthopade*, 29 (Suppl 1), 3.

Price, A.J., Longino, D., Rees, J., Rout, R., Pandit, H., Javaid, K., Arden, N., Cooper, C., Carr, A.J., Dodd, C.A.F., Murray, D.W., Beard, D.J., 2010. Are pain and function better measures of outcome than revision rates after TKR in the younger patient? *Knee*, 17, pp.196-99.

Price, A.J., Rees, J.L., Beard, D., Juszczak, E., Carter, S., White, S., de Steiger, R., Dodd, C.A.F., Gibbons, M., McLardy-Smith, P., Goodfellow, J.W., Murray, D.W., 2003. A mobile-bearing total knee prosthesis compared with a fixed-bearing prosthesis; a multicentre single blind randomised controlled trial. *J Bone Joint Surg*, 85-B(1), pp.62-67.

Price, D.D., Bush, .FM., Long, S., Harkins, S.W., 1994. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain*, 56(2), pp.217-26.

Putti, V., 1920. Arthroplasty of the knee joint. *J Orthop Surg*, 2, pp.530.

Ranawat, C.S., Flynn, W.F. Jr., Saddler, S., Hansraj, K.K., Maynard, M.J., 1993. Long-term results of the total condylar knee arthroplasty. A 15-year survivorship study. *Clin Orthop Relat Re*s, 286, pp.94-102.

Reimann, J., Brimah, K., Schroder, R., Wernig, A., Beauchamp, J.R., Partridge, T.A., 2004. Pax7 distribution in human skeletal muscle biopsies and myogenic tissue cultures. *Cell Tissue Res*, 315, pp: 233–42.

Relaix, F., Rocancour,t D., Mansouri, A., Buckingham, M., 2005. "A Pax3/Pax7-dependent population of skeletal muscle progenitor cells." *Nature*, 435 (7044), pp.898–9.

Riley, P.O., Schenkman, M.L., Mann, R.W., et al. 1991. Mechanics of a constrained chair-raise. *J Biomech*, 24, pp.77.

Renault, V., Thorne, L.E., Erikson, P.O., Butler-Browne, G., Mouly, V., 2002. Regenerative potential of human skeletal muscle during aging. *Aging Cell*, 1, pp.132-9.

Riddle, D.L., Stratford, P.W., Bowman, D.H., 2008. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trails: a systematic review. *Arthritis Rheum*, 59(6), pp.876-83.

Robertsson, O., Dunbar, M., Pehrsson, T., Knutson, K., Lidgren, L., 2000. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. *Acta Orthop Scand*, 71, pp.262–7.

Robertson, S., Frost, H., Doll, H., O'Connor, J.J., 1998. Leg extenspr power and quadriceps strength: an assessment of repearability in patients with osteoarthritic knees. *Clin Rehab*, 12, pp.120-5.

Rodrigues, A., and Schmalbruch, H., 1995. Satellite cells and myonuclei in long term denervated rat muscles. *Anat Rec*, 243, pp.430-37.

Roth, S.M., Ivey, F.M., Martel, G.F., Lemmer, J.T., Hurlbut, D.E., Siegel, E.L., Metter, E.J., Fleg, J.L., Fozard, J.L., Kostek, M.C., Wernick, D.M., Hurley, B.F., 2001. Muscle size responses to strength training in young and older men and women. *J Am Geriatr Soc*, 49-11, pp.1428-33.

Ryall, J.G., Schertzer, J.D., Lynch. G.S., 2008. Cellular and molecular mechanisms underlying agerelated skeletal muscle wasting and weakness. *Biogerontology*, 9, pp.213–28.

Sathasivam, S., Walker, P.S., 1994. Optimization of the bearing surface geometry of total knees. *J Biomech*, 27(3), pp.255-64.

Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., Rudnicki, M.A., 2000. Pax7 is required for the specification of myogenic satellite cells. *Cell*, 102, pp.777–86

Schubert, W., Zimmermann, K., Cramer, M., Starzinski-Powitz, A., 1989. Lymphocyte antigen Leu-19 as a molecular marker of regeneration in human skeletal muscle. *Proc Natl Acad Sci USA*, 86, pp;307–11.

Schultz, E., Gibson, M.C., and Champion, T., 1978. Satellite cells are mitotically quiescent in mature mouse muscle: an EM and radioautographic study. *J. Exp. Zool*. 206, pp.451-456.

Schultz, E., 1976. Fine structure of satellite cells in growing skeletal muscle. *Am J Anat*, 147, pp.49–70.

Shetty, A.A., Tindall. A., Ting, P., Heatley, F.W., 2003a. The evolution of total knee arthroplasty. Part I: introduction and first steps. *Current Orthopaedics* 17, pp.322-5.

Shetty, A.A.., Tindall, A., Ting, P., Heatley, F.W., 2003b. The evolution of total knee arthroplasty. Part II: the hinged knee replacement and the semi-constrained knee replacement. *Current Orthopaedics*, 17, 403-7.

Shetty, A.A., Tindall, A., Ting, P., Heatley, F.W., 2003c. The evolution of total knee arthroplasty. Part III: surface replacement. *Current Orthopaedics*, 17, pp.478-81.

Silva, M., Shepard, E.F., Jackson, W.O., Pratt, J.A., McClung, C.D., Schmalzried, T.P., 2003. Knee Strength after Total Knee Arthroplasty. *J Arthroplasty*, 18, pp.605-11.

Singh, A., Schmalzried, T., 1996. The extensor apparatus of the knee: decreased mechanical advantage following total knee replacement. *Orthop Trans*, 20, pp.178

Snijders, T., Verdijk, L.B., van Loon, L.J., 2009. The impact of sarcopenia and exercise training on skeletal muscle satellite cells. *Ageing Res Rev* 8, pp.328–338

Snow, M.H., 1978. An autoradiographic study of satellite cell differentiation into regenerating myotubes following transplantation of muscles in young rats. *Cell Tissue Res*, 186, pp.535–40.

Spangenberg, E., Booth, F., 2001. Invited Editorial on Myogenic satellite cells: Physiology to molecular biology. *J Appl Physiol*, 91(2), pp.533.

Stratford, P.W., and Kennedy, D.M., 2006. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemol*, 59, pp.160-7.

Studitsky, A.N., 1964. Free auto- and homografts of muscle tissue in experiments on animals. *Ann. N.Y. Acad. Sci*. 120, pp.789-801.

Su, F., Lai, K., Hong, W., 1998. Raising from a chair after total knee arthroplasty. *Clin Biomech*, 13, pp.176.

Terwee, C.B., Rienk, M.A., van der Slikke, R.M.A., van Lummel, R.C., Benink, R.J., Meijers, W.G.H., de Vet, H.C.W. 2006. Self-reported physical functioning was more influenced by pain than performance based physical functioning in knee-osteoarthritis patients. *J Clin Epidemol*, 59, pp.724-31.

The KAT trial group. 2009. The knee arthroplasty trial (KAT): design features, baseline characteristics, and two-year functional outcomes after alternative approaches to knee replacement. *J Bone Joint Surg [Am]*, 91, pp.134-41.

Timmons, N., 2008. NHS goes to the PROMS. BMJ, 336, pp.1464-1465. doi:10.1136/bmj.39618.627951.80 (Published 26 June 2008).

Tisselius, P., 1969. Studies on joint temperature, joint stiffness and muscle weakness in rheumatoid arthritis. *Acta Rheumatologica Scand*, 14(Suppl 14), 1.

Todd, R.C., Barnert, P., Day, W.H., 1979. ICLH arthroplasty of the knee 1968-1978. *J Bone Joint Surg [Br]*, 61-B, pp.339.

Valtonen, A., Poyhonen, T., Hinonen, A., Sipila, S., 2009. Muscel defecits persist after unilateral knee replacement and have implications for rehabilitation. *Phys Ther*, 89(10), pp1072-9.

Van Koff, M., Jensen, M., Karloy, P., 2000. Assessing Global Pain Severity by Self-Report in Clinical and Health Services Research Spine, 24, pp.3140-51.

Verdijk, L.B., Koopman, R., Schaart, G., Meijer, K., Savelberg, H.H., van Loon, L.J., 2007. Satellite cell content is specifically reduced in type II skeletal muscle fibers in the elderly. *Am J Physiol Endocrinol Metab*, 292, pp.E151–7.

Wagers, A.J., Conboy, I.M., 2005. Cellular and molecular signatures of muscle regeneration: current concepts and controversies in adult myogenesis. *Cell*, 122-5, pp.659-67.

Walldius, B., 1960. Arthroplasty of the knee using an endoprosthesis. Acta Orthop Scand, 30, pp.137.

Walldius, B., 1957. Arthroplasty of the knee using an endoprosthesis. Acta Orthop Scand (Suppl), 24.

Walldius, B., 1953. Arthroplasty of the knee using acrylic prosthesis. *Acta Orthop Scand*, 23, pp.121.

Walsh, M., Woodhouse, L.J., Thomas, S.G., Finch, E., 1998. Physical impairments and functional limitation: a comparison of individuals 1 year after total knee arthroplasty with control subjects. *Phys Ther*, 78, pp.248-58.

Wang, H., Simpson, K., Farrara, M., Chamnongkich, S., Kinsey, T., Mahoney, O.M., 2006. Biomechanical differences exhibited during sit-to-stand between total knee arthroplasty designs of varying radii. *J Arthroplasty*, 21, pp.1193-9.

Watkins, M.A., Riddle, D.L., Lamb, R.L., Personius, W.J., 1991. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Phys Ther*. 71(2), pp.90-6.

Whipple, R.H., Wolfson, L.I., Amerman, P.M., 1987. The relationship of knee and ankle weakness to falls in nursing home residents: an isokinetic study. *J Am Geriatric Soc*, 35, pp.13-20

Williamson, A., Hogarth, B., 2005. Pain: a review of three commonly used pain rating scales. *J Clin Nursing*, 14(7), pp.798-804.

Witvrouw, E., Victor, J., Bellemans, J., Rock, B., van Lummel, R., van der Slikke, R., Verdonk, R., 2002. A correlation study of objective functionality and WOMAC in total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*, 10, pp.347-51.

Wood, L., Peat, G., Thomas, E., Hay, E.M., Sim, J., 2008. Associations Between Physical Examination and Self-Reported Physical Function in Older Community-Dwelling Adults With Knee Pain. *Phys Ther*, 88, pp.33-42.

Wright, R.J., Sledge, C.B., Poss, R., Ewald, F.C., Walsh, M.E., Lingard, E.A., 2004. Patient-reported outcome and survivorship after Kinemax total knee arthroplasty. *J Bone Joint Surg [Am]*, 86-A, pp.2464–70.

Wylde, V., Dieppe, P., Hewlet, S., Learmouth, I.D., 2007. Total knee replacement: Is it really an effective procedure for all? *Knee*, 14. pp.417-23

Yoshida, Y., Mizner, R.L., Ramsey, D.K., Snyder-Mackler L., 2007. Examining outcomes from total knee arthroplasty and the relationship between quadriceps strength and knee function over time. *Clin Biomech*, 23, pp.320-8.

Zammit, P., 2008. All muscle satellite cells are equal, but are some more equal than others? *J Cell Sci*, 121, pp.2975-82.

Zammit, P.S., Partridge, T.A., and Yablonka-Reuveni, Z., 2006. The skeletal muscle satellite cell: the stem cell that came in from the cold. *J. Histochem. Cytochem*. 54, pp.1177-91.